(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 July 2002 (11.07.2002)

PCT

(10) International Publication Number
WO 02/053589 A2

(51) International Patent Classification[7]: C07K 14/415, 16/16, C12Q 1/68, G01N 33/50, C12N 15/11, 15/62, 5/10

(21) International Application Number: PCT/EP02/00073

(22) International Filing Date: 7 January 2002 (07.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/259,890 5 January 2001 (05.01.2001) US

(71) Applicant *(for all designated States except US)*: **CROPDESIGN N.V.** [BE/BE]; Technologiepark 3, B-9052 Zwijnaarde (BE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DE VEYLDER, Lieven** [BE/BE]; Josef Boddaertstraat 23, B-9031 Drongen (BE). **BOUDOLF, Veronique, Katelijne, Cecile, Kristien** [BE/BE]; Hettingen 47, B-9860 Oosterzele (BE). **INZE, Dirk** [BE/BE]; Driesstraat 18, B-9310 Moorsel-Aalst (BE). **FRANKARD, Valérie, Marie-Noëlle** [BE/BE]; Rue du Ham 40, B-1180 Bruxelles (BE). **TERRAS, Franky** [BE/BE]; Rijsenbergstraat 146L, B-9000 Gent (BE).

(74) Agents: **DE CLERCQ, Ann** et al.; De Clercq, Brants & Partners, E. Gevaertdreef 10 a, B-9830 Sint-Martens-Latem (BE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/053589 A2

(54) Title: NUCLEIC ACID MOLECULES ENCODING DIM INTERACTORS AND USES THEREFOR

(57) Abstract: The invention provides isolated nucleic acids molecules, designated DIM1 interacting molecules (DIMIC) molecules, which encode novel cell cycle associated polypeptides. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing DIMIC nucleic acid molecules, host cells into which the expression vectors have been introduced, and transgenic plants in which a DIMIC gene has been introduced or disrupted. The invention still further provides isolated DIMIC proteins, fusion proteins, antigenic peptides and anti-DIMIC antibodies. Agricultural, and screening methods utilizing molecules and compositions of the invention are also provided. The invention further provides methods for modulating overall growth and yield in a plant, enhancing stress tolerance of a plant, conferring enhanced resistance to pathogens to a plant, modulating pre-mRNA splicing in a plant cell, or modulating vesicle transport/processing in a plant cell, comprising introducing into the plant or plant cell a DIMIC molecule or DIMIC modulator, alone or in combination with a DIM1 molecule.

NUCLEIC ACID MOLECULES ENCODING DIM INTERACTORS AND USES THEREFOR

Background of the Invention

When eukaryotic cells and, thus, also plant cells divide they go through a highly ordered sequence of events collectively termed as the "cell cycle." Briefly, DNA replication or synthesis (S) and mitotic segregation of the chromosomes (M) occur with intervening gap phases (G1 and G2) and the phases follow the sequence G1-S-G2-M. Cell division is completed after cytokinesis, the last step of the M-phase. Cells that have exited the cell cycle and have become quiescent are said to be in the G0 phase. Cells at the G0 stage can be stimulated to re-enter the cell cycle at the G1 phase.

The transition between the different phases of the cell cycle is basically driven by the sequential activation/inactivation of a kinase, termed cyclin-dependent kinase or Cdk (*e.g.*, Cdc2 in *Schizosaccharomyces pombe* and in plants, Cdc18 in *Saccharomyces cerevisiae*), by different agonists. Also required for kinase activation are proteins called cyclins which are important for targeting the kinase activity to a given subset of substrate(s). Other factors regulating Cdk activity include Cdk inhibitors (CKIs or ICKs, Kips, Cips, Inks), Cdk activating kinase (CAK), Cdk phosphatase (Cdc25) and Cdk subunit (CKS) (Mironov *et al.* (1999) *Plant Cell* 11, 509-522 and Reed (1996) *Prog Cell Cycle Res* 2, 15-27 for reviews).

The *dim1+* gene was first isolated in the yeast *S. pombe* during a screen for second site mutations capable of reducing the restrictive temperature of the fission yeast mutant cdc2-D127N (Berry and Gould (1997) *J Cell Biol* 137, 1337-1354). When shifted to restrictive temperature, *dim1*-35 mutant cells arrest before entry into mitosis or proceed through mitosis in the absence of nuclear division, demonstrating an uncoupling of proper DNA segregation from other cell cycle events. Deletion of *dim1* from the *S. pombe* genome produces a lethal G2 arrest phenotype. Lethality is rescued by overexpression of the mouse *dim1* homologue, *mdim1*. Likewise, deletion of the *S. cerevisiae dim1* homologue, *DIB1*, is lethal. Both *mdim1* and *dim1+* are capable of rescuing lethality in the *dib1*::HIS3 mutant. *DIB1* was also termed *CDH1* ('*Saccharomyces cerevisiae* DIM1 homologue) by Berry and Gould (1997), *J Cell Biol* 137, 1337-1354. This alternative terminology is, however, confusing because of the existence of another *S. cerevisiae CDH1* gene which is a CDC20/Fizzy-homolog.

Although *dim1*-35 displays no striking genetic interactions with various other G2/M or mitotic mutants, *dim1*-35 cells incubated at a restrictive temperature arrest with low histone H1 kinase activity. Moreover, *dim1*-35 displays sensitivity to the microtubule

destabilizing drug, thiabendazole (TBZ). Those results suggest that Dim1p plays a fundamental, evolutionarily conserved role in the entry of cells into mitosis and in chromosome segregation during mitosis. Based on TBZ sensitivity and failed chromosome segregation in dim1-35, it can also be presumed that Dim1p may play a role in mitotic spindle formation and/or function.

To further understand dim1p function, Berry *et al.* (1999), *Mol Cell Biol* 19, 2535-2546, undertook a synthetic lethal screen with the temperature- sensitive *dim1*-35 mutant and isolated *lid* (for lethal in *dim1*-35) mutants. One of the *lid* mutants is the temperature sensitive *lid1*-6 mutant. At the restrictive temperature of 36°C, *lid1*-6 mutant cells arrest with a "cut" phenotype similar to that of *cut4* and *cut9* mutants, that are components of the anaphase promoting complex/cyclosome (APC/C; Tyers and Jorgensen (2000) *Curr Opin Genet Dev* 10, 54-64). The *S. pombe* genes *cut4* and *cut9* have known homologues in *S. cerevisiae* (*apc1* and *cdc16*, respectively) and at least for *cut9*, a metazoan homologue exists (*APC6*). An epitope tagged version of lid1p is a component of a multiprotein ~20S complex; the presence of lid1p in this complex depends upon the presence of a functional *cut9*$^{+}$. Lid1p-myc coimmunoprecipitates with several other proteins, including the APC/C members cut9p and nuc2p, and the presence of cut9p in a 20S complex depends upon the activity of *lid1*$^{+}$. Further, *lid1*$^{+}$ function is required for the multi-ubiquitination of cut2p, an anaphase-promoting complex (APC/C) target. Thus, lid1p is a component of the *S. pombe* APC/C. In *dim1* mutants, the abundance of lid1p and the APC/C complex decline significantly, and the ubiquitination of an APC/C target is abolished. These data suggest that at least one role of dim1p is to maintain or establish the steady state level of the APC/C.

Human HEF1 is a member of a family of multidomain docking proteins implicated in the regulation of cell adhesion. Expression of HEF1 is cell cycle regulated. The differentially phosphorylated p105^{HEF1} and p115^{HEF1} proteins are produced upon induction of cell growth and accumulate predominantly in the cytoplasm and to focal adhesions. The p55^{HEF1} protein, however, appears at mitosis as the result of processing by a caspase and localizes to the mitotic spindle. The human homolog of the *S. pombe* dim1p protein, hDIM1, was identified in a two-hybrid library screen as an interactor of p55^{HEF1} (Law *et al.* (1998) *Mol Cell Biol* 18, 3540-3551).

Upon purification of the *S. cerevisiae* U4/U6·U5 small nuclear ribonucleoprotein (snRNP) particle and subsequent identification of its protein constituents, it was found that Dib1, the yeast homolog of the *S. pombe* dim1p and human hDIM1, is an integral

component of the U4/U6·U5 snRNP. It was further argued that the previously described cell cycle defects associated with dim1p mutations may be secondary effects arising from defective pre-mRNA splicing (Stevens and Abelson (1999) *Proc Natl Acad Sci USA* 96, 7226-7231). The identification of an FKBP-type peptidyl-prolyl *cis-trans* isomerase motif in the different DIM1-homologues (Zhang *et al.* (1999) *Physiol Genomics* 1, 109-118) further strengthens the possible involvement of the DIM1 protein in pre-mRNA splicing as prolyl isomerases contribute to this process (Teigelkamp *et al.* (1998) *RNA* 4, 127-141).

At least the human DIM1 protein belongs to the superfamily of proteins adopting a thioredoxin fold. However, none of the DIM1 members contain the CGPC amino acid motif which is required for thioredoxin activity. Therefore, DIM1 proteins are most likely not active as thioredoxins (Zhang *et al.* (1999) *Physiol Genomics* 1, 109-118). Known dominant-negative mutants include C-terminal truncated DIM1 (deletion of the C-terminal 13 or 14 amino acids; Zhang *et al.* (1999) *Physiol Genomics* 1, 109-118). A temperature-sensitive mutant , *dim1*-35, is known in *S. pombe*. In *dim1*-35, a single amino acid is changed relative to wild-type *dim1*, namely the wild-type glycine at position 126 that is changed into an aspartate in *dim1*-35 (Berry and Gould (1997) *J Cell Biol* 137, 1337-1354). The *S. cerevisiae* Dib1 protein should not be confused with the *S. cerevisiae* Dim1 protein. Whereas Dib1 is the yeast homolog of the *S. pombe* dim1p and the human hDIM1, the yeast Dim1 protein is an 18s rRNA dimethylase (Lafontaine *et al* (1994) *J. Mol. Biol.* 241, 492-497). The Arabidopsis homolog of the yeast *DIM1* rRNA methylase gene is known as *PFC1* (*PALEFACE1*; Tokuhisa *et al.* (1998) *Plant Cell* 10, 699-711).

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel plant nucleic acid molecules and polypeptides encoded by such nucleic acid molecules, referred to herein as "DIM1-interacting molecules" or "DIMIC." The DIMIC nucleic acid and polypeptide molecules of the present invention are useful as modulating agents in regulating cell cycle progression in, for example, plants. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding DIMIC polypeptides, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of DIMIC-encoding nucleic acids.

According to a first embodiment the present invention relates to an isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule comprising the nucleotide sequence as given in any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof,

(b) a nucleic acid molecule comprising the RNA sequence corresponding to any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof,

(c) a nucleic acid molecule specifically hybridizing with the nucleotide sequence as defined in (a) or (b),

(d) a nucleic acid molecule which is at least 60% identical to the nucleotide sequence as given in any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof,

(e) a nucleic acid molecule encoding a protein comprising an amino acid sequence as given in any of SEQ ID NOs 49 to 53 or 95,

(f) a nucleic acid molecule encoding a protein comprising at least one or at least two or at least three of the amino acid sequences represented in SEQ ID NOs 55, 56 or 96,

(g) a nucleic acid molecule encoding a protein comprising an amino acid sequence which is at least 42 % identical to the amino acid sequence as given in SEQ ID NO 50,

(h) a nucleic acid molecule encoding a protein comprising at least one or at least two, or three, or four, or five, or six, or seven, or eight, or nine, or ten, or eleven, or twelve, or thirteen, or fourteen of the amino acid sequences represented in SEQ ID NOs 59 to 63, or 97 to 105,

(i) a nucleic acid molecule encoding a protein comprising at least one or at least two of the amino acid sequences represented in SEQ ID NOs 64 to 69, 106, 107 or 111,

(j) a nucleic acid molecule encoding a protein comprising at least one or two or three of the amino acid sequences represented in any of SEQ ID NOs 108, 109 or 110,

(k) a nucleic acid molecule encoding a protein comprising an amino acid sequence which is at least 50 % identical to the amino acid sequence as given in any of SEQ ID NOs 49, 50, 51, 52, 53 or 95,

(l) a nucleic acid molecule which is degenerated to a nucleic acid as defined in any of (a) to (k) as a result of the genetic code,

(m) a nucleic acid molecule which is diverging from a nucleic acid as defined in any of (a) to (k) as a result of differences in codon usage between organisms,

(n) a nucleic acid molecule which is diverging from a nucleic acid as defined in any of (a) to (k) as a result of differences between alleles, and

(o) a nucleic acid molecule as defined in any one of (a) to (n) characterized in that said nucleic acid is DNA, cDNA, genomic DNA or synthetic DNA,

characterized in that said nucleic acid molecule encodes a DIM1-interacting molecule (DIMIC molecule), or a homologue or a derivative thereof and further provided that said nucleic acid is not one of the nucleic acids as deposited under the GenBank Accession numbers AC004261, AC008148, AB023039 or AC007583.

According to another embodiment, the invention relates to an isolated nucleic acid molecule encoding an immunologically active and/or functional fragment of a DIM1-interacting molecule encoded by a nucleic acid of claim 1, or an immunologically active and/or functional fragment of a homologue or a derivative of such a DIM1-interacting molecule, provided that said nucleic acid is not one of the nucleic acids as deposited under the GenBank Accession number T3K9.20 or T3K9.21. For instance, said isolated nucleic acid molecule are selected from the group consisting of consisting of:

(a) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 49, wherein said fragment comprises at least one or two or three of the sequences as represented in any of SEQ ID NOs 55, 56, or 96,

(b) a nucleic acid molecule encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 49, wherein, said fragment comprises at least 326 contiguous amino acid residues of the amino acid sequence of SEQ ID NO 49,

(c) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 50, wherein said fragment comprises at least one, or two, or three, or four, or five, or six, or seven, or eight, or nine, or ten, or eleven, or twelve, or thirteen, or fourteen of the sequences as represented in any of SEQ ID NOs 59, 60, 61, 62, 63, 97, 98, 99, 100, 101, 102, 103, 104, or 105,

(d) a nucleic acid encoding a functional fragment of polypeptide comprising the amino acid sequence of SEQ ID NO 51, wherein said fragment comprises at least one, or two, or three, or four, or five, or six, or seven, or eight, or nine of the sequences as represented in any of SEQ ID NOs 64, 65, 66, 67, 68, 69, 106, 107 or 111,

(e) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 52, wherein said fragment comprises at least one or two of the sequences as represented in SEQ ID NO 108 or 110,

(f) a nucleic acid molecule encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 53, wherein said fragment comprises at least one or two of the sequences as represented in SEQ ID NO 109 or 110,

(g) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 95, wherein said fragment comprises at least one or two of the sequences as represented in SEQ ID NO 109 or 110, and

(h) a nucleic acid molecule encoding a functional fragment of a polypeptide comprising the amino acid sequence of any SEQ ID NOs 52, 53 or 95, wherein the fragment comprises at least 178 contiguous amino acid residues of any of the amino acid sequences of SEQ ID NOs 52, 53 or 95.

In one embodiment, a DIMIC nucleic acid molecule of the invention is at least 50%, 55%, 57%, 60%, 62%, 65%, 67%, 70%, 72%, 75%, 77%, 80%, 82%, 85%, 87%, 90%, 92%, 95%, 97%, 98%, 99% or more identical to the nucleotide sequence (*e.g.*, to the entire length of the nucleotide sequence) of SEQ ID NOs 35-48 or 94, or a complement thereof.

In a preferred embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown in SEQ ID NOs 35-48 or 94, or a complement thereof. In another preferred embodiment, an isolated nucleic acid molecule of the invention encodes the amino acid sequence of a plant DIMIC polypeptide.

According to a further embodiment the present invention relates to an isolated nucleic acid molecule consisting of the nucleotide sequence as given in any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof.

Another embodiment of the invention features nucleic acid molecules, preferably DIMIC nucleic acid molecules, which specifically detect DIMIC nucleic acid molecules relative to nucleic acid molecules encoding non-DIMIC polypeptides. For example, in one embodiment, such a nucleic acid molecule is at least 15, 20, 25, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 532, 550, 600, 650, 700, 750, 800, 850, 900, 950, 976 or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid

molecule comprising the nucleotide sequence shown in SEQ ID NOs 35-48 or 94, or a complement thereof.

In another embodiment, the present invention features fragments of the nucleic acid molecule of SEQ ID NOs 35, 36, or 37, wherein the fragments do not comprise nucleotides 1-975 (SEQ ID NO 85), nucleotides 1087-1236 (SEQ ID NO 86), nucleotides 1237-1326 (SEQ ID NO 87), or nucleotides 1330-1599 (SEQ ID NO 88) of SEQ ID NOs 35, 36, or 37.

In another embodiment, the present invention features fragments of the nucleic acid molecule of SEQ ID NOs 44, 45, 46, 47, or 48, wherein the fragments do not comprise nucleotides 1-531 of SEQ ID NOs 44, 45, 46, 47, or 48 (SEQ ID NO 91), nucleotides 643-948 of SEQ ID NO 45 or 47 (SEQ ID NO 92), or nucleotides 646-810 of SEQ ID NO 46 or 48 (SEQ ID NO 93).

In other preferred embodiments, the nucleic acid molecule encodes a naturally occurring allelic variant of a plant DIMIC polypeptide, wherein the nucleic acid molecule hybridizes to the nucleic acid molecule of SEQ ID NOs 35-48 or 94 under stringent conditions.

Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to a DIMIC nucleic acid molecule, *e.g.*, the coding strand of a DIMIC nucleic acid molecule.

Therefore the invention also relates to an antisense nucleic acid molecule corresponding to at least one of the DIMIC nucleic acids as described earlier.

The invention also relates to an isolated nucleic acid molecule comprising at least one of the DIMIC nucleic acids as described earlier and a nucleotide sequence encoding a heterologous polypeptide. The invention also relates to a polypeptide encodable by such an isolated nucleic acid.

The invention also relates to a nucleic acid molecule of at least 15 contiguous nucleotides in length specifically hybridizing with or specifically amplifying DIMIC nucleic acids as described earlier.

Another aspect of the invention provides a vector comprising a DIMIC nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector.

The invention therefore relates to a vector comprising any of the nucleic acid molecules of the invention and as described earlier. The invention further relates to an expression vector wherein said nucleic acid sequence of the invention is operably linked to one or more control sequences allowing the expression of said sequence in prokaryotic and/or eukaryotic host cells.

In another embodiment, the invention provides a host cell containing a vector of the invention. Therefore the invention relates to host cell comprising a nucleic acid molecule of the invention or a vector as described above, for instance a host cell chosen from a bacterial, insect, fungal, yeast, plant or animal cell.

The invention also provides a method for producing a DIMIC polypeptide, by culturing in a suitable medium a host cell of the invention, *e.g.*, a plant host cell such as a host monocot plant cell (*e.g.*, rice, wheat or corn) or a dicot host cell (*e.g.*, Arabidopsis thaliana, oilseed rape, or soybeans) containing a recombinant expression vector, such that the polypeptide is produced.

The invention thus relates to a method for producing a polypeptide comprising culturing a host cell as described above under conditions allowing the expression of the polypeptide and recovering the produced polypeptide from the culture.

Another aspect of this invention features isolated or recombinant DIMIC polypeptides.

The invention relates to an isolated polypeptide encodable by any of the nucleic acids of the invention, or a homologue or a derivative thereof, or an immunologically active and/or functional fragment thereof. For instance, the invention relates to a polypeptide having an amino acid sequence as given in any of SEQ ID NOs 49 to 53 or 95, or a homologue or a derivative thereof, or an immunologically active and/or functional fragment thereof.

In one embodiment, an isolated DIMIC polypeptide has one or more of the following domains: a "WW or WWP domain", a "non-classical C_2-domain", a "FAB1 activation loop", a "DIMIC5 internal repeat domain", a "DIMIC7 internal repeat domain", a "DIMIC26 internal repeat domain", a "DIMIC26 di-amino acid motif", a "thioredoxin-like domain" and/or a "PEST sequence."

In a preferred embodiment, a DIMIC polypeptide includes at least one or more of the following domains: a "WW or WWP domain", a "non-classical C_2-domain", a "FAB1 activation loop", a "DIMIC5 internal repeat domain", a "DIMIC7 internal repeat domain", a "DIMIC26 internal repeat domain", a "DIMIC26 di-amino acid motif", a "thioredoxin-like domain" and/or a "PEST sequence", and has an amino acid sequence at least about 50%, 52%, 55%, 57%, 60%, 62%, 65%, 67%, 70%, 72%, 75%, 77%, 80%, 82%, 85%, 87%, 90%, 92%, 95%, 97%, 98%, 99% or more identical to the amino acid sequence of SEQ ID NOs 49-53 or 95.

In another preferred embodiment, a DIMIC polypeptide includes at least one or more of the following domains: a "WW or WWP domain", a "non-classical C_2-domain", a

"FAB1 activation loop", a "DIMIC5 internal repeat domain", a "DIMIC7 internal repeat domain", a "DIMIC26 internal repeat domain", a "DIMIC26 di-amino acid motif", a "thioredoxin-like domain" and/or a "PEST sequence" and has a DIMIC activity (as described herein).

In yet another preferred embodiment, a DIMIC polypeptide includes one or more of the following domains: a "WW or WWP domain", a "non-classical C_2-domain", a "FAB1 activation loop", a "DIMIC5 internal repeat domain", a "DIMIC7 internal repeat domain", a "DIMIC26 internal repeat domain", a "DIMIC26 di-amino acid motif", a "thioredoxin-like domain" and/or a "PEST sequence" and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOs 35-48 or 94.

In another embodiment, the invention features fragments of the polypeptide having the amino acid sequence of SEQ ID NOs 49-53 or 95, wherein the fragment comprises at least 178, 200, 250, 300, 326, 350, or more amino acids (*e.g.*, contiguous amino acids) of the amino acid sequence of SEQ ID NOs 49-53 or 95. In another embodiment, a DIMIC polypeptide has the amino acid sequence of SEQ ID NOs 49-53 or 95.

In a further embodiment, the invention features fragments of the polypeptide having the amino acid sequence of SEQ ID NO 49, wherein the fragments do not comprise amino acid residues 1-325 (SEQ ID NO 81), amino acid residues 363-412 (SEQ ID NO 82), amino acid residues 413-442 (SEQ ID NO 83), or amino acid residues 444-463 (SEQ ID NO 84) of SEQ ID NO 49.

In another embodiment, the invention features fragments of the polypeptide having the amino acid sequence of SEQ ID NO 52 or 53, wherein the fragments do not comprise amino acid residues 1-177 of SEQ ID NO 52 (SEQ ID NO 89), amino acid residues 215-268 of SEQ ID NO 52, or amino acid residues 216-269 of SEQ ID NO 53 (SEQ ID NO 90).

In another embodiment, the invention features a DIMIC protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 50%, 52%, 55%, 57%, 60%, 62%, 65%, 67%, 70%, 72%, 75%, 77%, 80%, 82%, 85%, 87%, 90%, 92%, 95%, 97%, 98%, 99% or more identical to a nucleotide sequence of SEQ ID NOs 35-48 or 95, or a complement thereof. This invention further features a DIMIC polypeptide, which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid

molecule comprising the nucleotide sequence of SEQ ID NOs 35-48 or 95, or a complement thereof.

In another embodiment the invention provides transgenic plants (*e.g.*, monocot or dicot plants) containing an isolated nucleic acid molecule of the present invention. For example, the invention provides transgenic plants containing a recombinant expression cassette including a plant promoter operably linked to an isolated nucleic acid molecule of the present invention. The present invention also provides transgenic seed from the transgenic plants. In another embodiment the invention provides methods of modulating, in a transgenic plant, the expression of the nucleic acids of the invention.

The invention thus relates to a method for the production of altered plant cells, plant tissues or plants comprising the introduction of a polypeptide as defined earlier directly into said plant cell or tissue or in an organ of said plant.

The invention also relates to a method for effecting the expression of a polypeptide as defined earlier in plant cells, tissues or plants comprising the introduction of any of the nucleic acid molecules of the invention operably linked to one or more control sequences or a vector of the invention stably into the genome of a plant cell.

The invention also relates to a method for the production of transgenic plant cells, plant tissues or plants comprising the introduction of a nucleic acid of the invention in an expressible format or a vector of the invention in said plant cell, plant tissue or plant.

The invention also relates to a method as described above further comprising regenerating a plant from said plant cell.

The invention further relates to a transgenic plant cell comprising any of the nucleic acids of the invention which is operably linked to regulatory elements allowing transcription and/or expression of said nucleic acid in plant cells or a transgenic plant cell obtainable by any of the methods described above. The invention relates to said transgenic plant cell wherein said nucleic acid is stably integrated into the genome of said plant cell.

The invention also relates to a transgenic plant or plant tissue comprising transgenic plant cells as described above or a transgenic plant obtainable by the method described above. The invention also relates to a harvestable part of said transgenic plant, for instance a harvestable part which is selected from the group consisting of seeds, leaves, fruits, stem cultures, rhizomes and bulbs. The invention also relates to the progeny derived from any of the transgenic plants or plant parts described above.

The proteins of the present invention or portions thereof, *e.g.*, biologically active portions thereof, can be operatively linked to a non-DIMIC polypeptide (*e.g.*, heterologous amino acid sequences) to form fusion proteins.

The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind polypeptide of the invention, preferably DIMIC polypeptide.

The invention thus relates to an antibody specifically recognizing a polypeptide of the invention or a specific epitope of said polypeptide.

In addition, the DIMIC polypeptide or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting the presence of a DIMIC nucleic acid molecule or polypeptide in a biological sample by contacting the biological sample with an agent capable of detecting a DIMIC nucleic acid molecule or polypeptide such that the presence of a DIMIC nucleic acid molecule or polypeptide is detected in the biological sample.

Therefore, the invention further relates to a method for detecting the presence of a polypeptide of the invention in a sample comprising:

(a) contacting the sample with a compound which selectively binds to said polypeptide, for instance an antibody; and

(b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of the invention in the sample.

The invention also relates to a method for detecting the presence of any of the nucleic acid molecules of the invention in a sample comprising:

(a) contacting the sample with a nucleic acid probe or primer as described earlier which selectively hybridizes to or amplifies one of the nucleic acid molecules of the invention, and

(b) determining whether said nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of said one nucleic acid molecule in the sample.

The invention further relates to the method described above, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

The invention further relates to a diagnostic kit comprising at least one of the nucleic acid molecules of the invention, at least one of the polypeptides of the invention,

at least one of the antibodies described above, at leat one of the compounds obtainable by any of the methods described further.

In another aspect, the present invention provides a method for detecting the presence of DIMIC activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of DIMIC activity such that the presence of DIMIC activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating DIMIC activity comprising contacting a cell capable of expressing DIMIC with an agent that modulates DIMIC activity such that DIMIC activity in the cell is modulated. In one embodiment, the agent inhibits DIMIC activity. In another embodiment, the agent stimulates DIMIC activity. In one embodiment, the agent is an antibody that specifically binds to a DIMIC polypeptide. In another embodiment, the agent modulates expression of DIMIC by modulating transcription of a DIMIC gene or translation of a DIMIC mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of a DIMIC mRNA or a DIMIC gene.

In one embodiment, the methods of the present invention are used to increase crop yield, improve the growth characteristics of a plant (such as growth rate or size of specific tissues or organs in the plant), modify the architecture or morphology of a plant, improve tolerance to environmental stress conditions (such as drought, salt, temperature, nutrient or deprivation), or improve tolerance to plant pathogens (*e.g.*, pathogens that abuse the cell cycle) by modulating DIMIC activity in a cell. In one embodiment, the DIMIC activity is modulated by modulating the expression of a DIMIC nucleic acid molecule. In yet another embodiment, the DIMIC activity is modulated by modulating the activity of a DIMIC polypeptide.

Modulators of the expression of DIMIC nucleic acids or DIMIC activity include, for example, a DIMIC nucleic acid such as an antisense version of a DIMIC nucleic acid molecule or a DIMIC polypeptide molecule. Other DIMIC modulators comprise antibodies to DIM1 or DIMIC molecules, small molecular weight compounds interacting with or modulating the activity of DIM1 or DIMIC molecules, ribozymes and the like.

Modulation of DIMIC activity can be achieved for instance by introducing a DIMIC nucleic acid molecule in a cell. This may lead to overexpression of the exogenous DIMIC molecule in said cell. Alternatively, this may lead to downregulation of expression of the endogenous DIMIC molecule in the cell, a phenomenon known under the term "silencing".

One example of such a DIMIC modulator is a nucleic acid molecule comprising at least part of the nucleotide sequence of a DIMIC molecule, for instance as represented in

any of SEQ ID NOs 35 to 48 or 94, and at least part of the corresponding antisense version of said part, seperated by at least a short stretch of nucleotides, in an inverted repeat confirmation.

Another example of such a DIMIC modulator is a fragment of said DIMIC polypeptide that, for instance, contains a destruction box, for instance a PEST sequence, which saturates the specific proteolytic machinery of the plant cell so that the endogenous polypeptide can survive longer in the plant cell. Other fragments of DIMIC polypeptides which can be used herein as a DIMIC modulator (or to modulate the activity of DIMIC molecules) are described further and comprise the specific polypeptide fragments of the DIMIC molecules of the invention (for instance SEQ ID NOs 81, 82, 83, 84, 89, 90, 96, 97, 108, 109) FKBP domains (for instance SEQ ID NO 54), WW or WWP domains (for instance SEQ ID NO 55), Non-classical C_2 domains (for instance SEQ ID NO 56), DIMIC5 internal repeat domains (for instance SEQ ID NO 57), FAB1 activation loops (for instance SEQ ID NO 58), DIMIC7 internal repeat domains (for instance any of SEQ ID NOs 59 to 63), DIMIC26 internal repeat domains (for instance any of SEQ ID NOs 64 to 69), DIMIC 26 di-amino acid motifs, thioredocin-like domains, PEST sequences (for instance any of SEQ ID NOs 98 to 105 or 107 or 110) and PHD fingers (for instance SEQ ID NO 111).

According to yet another embodiment the invention relates to a method for modulating the growth of a plant, comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the growth of the plant, thereby modulating the growth of the plant.

It should be understood herein that the DIM1 interacting molecule as used in the context for use in any of the methods described herein, comprises any DIM1 interacting molecule from prokaryotic or eukaryotic origin. In interesting embodiments, a plant DIM1 interacting molecule is used, in other interesting embodiments, at least one of the DIM1 interacting molecules identified herein is used.

The invention also relates to a method for modulating the cell cycle in a plant, comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the cell cycle in the plant, thereby modulating the cell cycle in the plant.

The invention further relates to a method for enhancing overall growth and yield of a plant comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the growth of the plant, thereby enhancing overall growth and yield of said plant.

The invention also relates to a method for increasing yield of a plant comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the growth of the plant, thereby increasing yield of said plant.

The present invention also relates to a method for enhancing stress tolerance, for instance osmotolerance or temperature tolerance, in a plant comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the growth of the plant, thereby enhancing stress tolerance of said plant. Also according to the invention is the use of a stress inducible promoter herein, to drive the expression of the DIMIC molecule or DIMIC modulator, with the aim to produce the osmoprotectant as it is necessary.

As used herein, "stress tolerance" refers to the capacity to grow and produce biomass during stress, the capacity to reinitiate growth and biomass production after stress, and the capacity to survive stress. The term "stress tolerance" also covers the capacity of the plant to undergo its developmental program during stress similarly to under non-stressed conditions, e.g. to switch from dormancy to germination and from vegetative to reproductive phase under stressed conditions similarly as under non-stressed conditions. Methodologies to determine plant growth or response to stress include, but are not limited to height measurements, leaf area, plant water relations, ability to flower, ability to generate progeny and yield or any other methodology known to those skilled in the art.

The expression "stress tolerance" as used herein preferably relates to tolerance against osmotic stress, caused by salt or drought and/or temperature stress, caused by cold, chilling and freezing stress.

The invention also relates to a method for conferring enhanced resistance to pathogens of a plant comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to confer resistance to pathogens of the plant, thereby conferring enhanced resistance to pathogens of said plant.

The invention further relates to any of the above described methods wherein at least one nucleic acid encoding a plant DIM1 interacting (DIMIC) molecule, a homologue or a derivative thereof or an enzymatically active fragment thereof is expressed in specific cells or tissues of said plant.

The invention further relates to the above method furhter comprising stably integrating into the genome of said plant or in specific plant cells or tissues of said plant

at least one expressible nucleic acid encoding a DIM1 interacting (DIMIC) molecule, a homologue or a derivative thereof or an enzymatically active fragment thereof

The invention further relates to any of the above methods wherein said expression of said nucleic acid leads to overexpresion of a DIM1 interacting (DIMIC) molecule in said plant or alternatively wherein said expression of said nucleic acid leads to downregulation of expression of a DIM1 interacting (DIMIC) molecule.

The invention furhte relates to any of the methods as described above wherein said DIM1 interacting (DIMIC) molecule is selected from any of the following nucleic acids:

(a) a nucleic acid molecule comprising the nucleotide sequence as given in any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof,

(b) a nucleic acid molecule comprising the RNA sequence corresponding to any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof,

(c) a nucleic acid molecule specifically hybridizing with the nucleotide sequence as defined in (a) or (b),

(d) a nucleic acid molecule which is at least 60% identical to the nucleotide sequence as given in any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof,

(e) a nucleic acid molecule encoding a protein comprising an amino acid sequence as given in any of SEQ ID NOs 49 to 53 or 95,

(f) a nucleic acid molecule encoding a protein comprising at least one or at least two or at least three of the amino acid sequences represented in SEQ ID NOs 55, 56 or 96,

(g) a nucleic acid molecule encoding a protein comprising an amino acid sequence which is at least 42 % identical to the amino acid sequence as given in SEQ ID NO 50,

(h) a nucleic acid molecule encoding a protein comprising at least one or at least two, or three, or four, or five, or six, or seven, or eight, or nine, or ten, or eleven, or twelve, or thirteen, or fourteen of the amino acid sequences represented in SEQ ID NOs 59 to 63, or 97 to 105,

(i) a nucleic acid molecule encoding a protein comprising at least one or at least two, or three, or four, or five, or six, or seven, or eight, or nine of the amino acid sequences represented in SEQ ID NOs 64 to 69, 106, 107 or 111,

(j) a nucleic acid molecule encoding a protein comprising at least one, or two or three of the amino acid sequences represented in any of SEQ ID NOs 108, 109 or 110,

(k) a nucleic acid molecule encoding a protein comprising an amino acid sequence which is at least 50 % identical to the amino acid sequence as given in any of SEQ ID NOs 49, 50, 51, 52, 53 or 95,

(l) a nucleic acid molecule which is degenerated to a nucleic acid as defined in any of (a) to (k) as a result of the genetic code,

(m) a nucleic acid molecule which is diverging from a nucleic acid as defined in any of (a) to (k) as a result of differences in codon usage between organisms,

(n) a nucleic acid molecule which is diverging from a nucleic acid as defined in any of (a) to (k) as a result of differences between alleles, and

(o) a nucleic acid molecule as defined in any one of (a) to (n) characterized in that said nucleic acid is DNA, cDNA, genomic DNA or synthetic DNA,

or to an isolated nucleic acid molecule encoding an immunologically active and/or functional fragment of a DIM1-interacting molecule encoded by a nucleic acid of any of (a) to (o), or an immunologically active and/or functional fragment of a homologue or a derivative of such a DIM1-interacting molecule selected from one of the following:

(a) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 49, wherein said fragment comprises at least one of the sequences as represented in any of SEQ ID NOs 55, 56, or 96,

(b) a nucleic acid molecule encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 49, wherein, said fragment comprises at least 326 contiguous amino acid residues of the amino acid sequence of SEQ ID NO 49,

(c) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 50, wherein said fragment comprises at least one of or at least two, or three, or four, or five, or six, or seven, or eight, or nine, or ten, or eleven, or twelve, or thirteen, or fourteen of the sequences as represented in any of SEQ ID NOs 59, 60, 61, 62, 63, 97, 98, 99, 100, 101, 102, 103, 104, or 105,

(d) a nucleic acid encoding a functional fragment of polypeptide comprising the amino acid sequence of SEQ ID NO 51, wherein said fragment comprises at least one of the sequences as represented in any of SEQ ID NOs 64, 65, 66, 67, 68, 69, 106, 107 or 111,

(e) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 52, wherein said fragment comprises at least one of the sequences as represented in SEQ ID NO 108 or 110,

(f) a nucleic acid molecule encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 53, wherein said fragment comprises at least one of the sequences as represented in SEQ ID NO 109 or 110,

(g) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 95, wherein said fragment comprises at least one of the sequences as represented in SEQ ID NO 109 or 110, and

(h) a nucleic acid molecule encoding a functional fragment of a polypeptide comprising the amino acid sequence of any SEQ ID NOs 52, 53 or 95, wherein the fragment comprises at least 178 contiguous amino acid residues of any of the amino acid sequences of SEQ ID NOs 52, 53 or 95.

The invention further relates to any of the methods described above wherein the expression or activity of a nucleic acid encoding a plant DIM1 interacting (DIMIC) molecule or a homologue thereof is modulated by a DIMIC modulator, for instance a DIMIC modulator selected from the group consisting of any of the described antibodies, antisense molecules, ribozymes, or compounds obtainable by any of the methods described further.

The invention further relates to the method described above wherein said DIMIC modulator is capable of modulating DIMIC nucleic acid expression or wherein said DIMIC modulator is capable of modulating DIMIC polypeptide activity.

According to yet another embodiment the invention also relates to any of the methods described earlier comprising co-expression of a DIM1 interacting (DIMIC) molecule or a DIMIC modulator and a DIM1 molecule in said plant.

The present inventors have performed a two hybrid screening with the Arabidopsis thaliana DIM1 (AtDIM1) as a bait to define a number of plant interacting proteins with AtDIM1 in yeast cells, for instance the DIM1 interacting (DIMIC) molecules as described earlier. These physical interactions are evidence that they occur in plant cells. Therefore the DIMIC molecules are the preferred partners to coexpress with DIM1. A number of DIM1 molecules are described herein and are represented in SEQ ID NOs 1 to 34.

In one example, co-expression of DIM1 and DIMIC5 in plants is performed. An effect on pre-mRNA splicing is expected, such as a more rapid and efficient intron splicing. More messenger can be translated into proteins, with a direct effect on cell growth, and thereafter on cell cycle progression.

Other DIMIC molecules to be coexpressed with a DIM1 molecule comprise for instance any of the DIMIC molecules described earlier, or a functional fragment thereof.

The invention also extends to the use of homologues, orthologues, paralogues or derivatives of the DIMIC molecules described herein and to functional fragments thereof.

Futhermore the invention also relates to new two hybrid screening methods which can be performed with any of the DIMIC molecules of the invention, for instance DIMIC5, DIMIC7, DIMIC26 or DIMIC70A/B/C, for instance to identify other interacting proteins of the spliceosome. Combined expression of these new interactors with either DIM1 or DIMIC5 or both in transgenic plants is yet another example to promote enhanced growth in plants.

With "co-expression" is meant the expression or overexpression of two or more genes or proteins. The same or, more preferably, different promoters are used to confer expression of said genes or proteins.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a DIMIC polypeptide; (ii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a DIMIC polypeptide, wherein a wild-type form of the gene encodes a protein with a DIMIC activity.

In another aspect the invention provides methods for identifying a compound that binds to or modulates the activity of a DIMIC polypeptide, by providing an indicator composition comprising a DIMIC polypeptide having DIMIC activity, contacting the indicator composition with a test compound, and determining the effect of the test compound on DIMIC activity in the indicator composition to identify a compound that modulates the activity of a DIMIC polypeptide. The identified compounds may be used as herbicides or plant growth regulators.

According to yet another embodiment the invention relates to a method for identifying compounds or mixtures of compounds which specifically bind to a polypeptide of the invention, comprising the steps of

(a) combining a polypeptide of the invention or a cell expressing said polypeptide with said compound or mixtures of compounds under conditions suitable to allow complex formation, and,

(b) detecting complex formation, wherein the presence of a complex identifies a compound or mixture of compounds which specifically binds said polypeptide.

The invention further relates to a method as described above, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

(a) detection of binding by direct detection of test compound/polypeptide binding;

(b) detection of binding using a competition binding assay; and

(c) detection of binding using an assay for testing the activity of the DIM1- interacting molecule.

According to a further embodiment the invention relates to a method for identifying and obtaining compounds interacting with or modulating the activity of a polypeptide of the invention comprising the steps of:

(a) providing a two-hybrid system wherein a polypeptide of the invention and an interacting protein partner, preferably a DIM1 molecule are expressed,

(b) interacting said compound with the complex formed by the expressed polypeptides as defined in a), and,

(c) performing measurement of interaction of said compound with said polypeptide or the complex formed by the expressed polypeptides as defined in (a).

The invention further relates to a method for modulating the activity of a polypeptide of the invention comprising contacting a polypeptide of the invention or a host cell of the invention expressing said polypeptide with a compound which binds to the polypeptide or obtainable by any of the methods described above, in a sufficient concentration to modulate the activity of the polypeptide.

The invention further relates to a method for preparing a DIMIC modulator composition using a compound identifiable by any of the methods described above.

In another aspect, the present invention features methods for modulating pre-mRNA splicing in a cell, *e.g.*, a plant cell, by introducing into the cell a DIMIC modulator in an amount sufficient to modulate pre-mRNA splicing in the cell, thereby modulating pre-mRNA splicing in the cell. In one embodiment, the DIMIC modulator comprises the nucleotide sequence of SEQ ID NOs 35-48, or a fragment thereof. In another embodiment, the DIMIC modulator is a DIMIC polypeptide comprising the amino acid sequence of SEQ ID NOs 49-53, or a fragment thereof. In another aspect, the present

(b) detecting complex formation, wherein the presence of a complex identifies a compound or mixture of compounds which specifically binds said polypeptide.

The invention further relates to a method as described above, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

(a) detection of binding by direct detection of test compound/polypeptide binding;

(b) detection of binding using a competition binding assay; and

(c) detection of binding using an assay for testing the activity of the DIM1- interacting molecule.

According to a further embodiment the invention relates to a method for identifying and obtaining compounds interacting with or modulating the activity of a polypeptide of the invention comprising the steps of:

(a) providing a two-hybrid system wherein a polypeptide of the invention and an interacting protein partner, preferably a DIM1 molecule are expressed,

(b) interacting said compound with the complex formed by the expressed polypeptides as defined in a), and,

(c) performing measurement of interaction of said compound with said polypeptide or the complex formed by the expressed polypeptides as defined in (a).

The invention further relates to a method for modulating the activity of a polypeptide of the invention comprising contacting a polypeptide of the invention or a host cell of the invention expressing said polypeptide with a compound which binds to the polypeptide or obtainable by any of the methods described above, in a sufficient concentration to modulate the activity of the polypeptide.

The invention further relates to a method for preparing a DIMIC modulator composition using a compound identifiable by any of the methods described above.

In another aspect, the present invention features methods for modulating pre-mRNA splicing in a cell, *e.g.*, a plant cell, by introducing into the cell a DIMIC modulator in an amount sufficient to modulate pre-mRNA splicing in the cell, thereby modulating pre-mRNA splicing in the cell. In one embodiment, the DIMIC modulator comprises the nucleotide sequence of SEQ ID NOs 35-48, or a fragment thereof. In another embodiment, the DIMIC modulator is a DIMIC polypeptide comprising the amino acid sequence of SEQ ID NOs 49-53, or a fragment thereof. In another aspect, the present

invention features methods for modulating vesicle transport/processing in a cell, *e.g.*, a plant cell, by introducing into the cell a DIMIC modulator in an amount sufficient to modulate vesicle transport/processing in the cell, thereby modulating vesicle transport/processing in the cell. In one embodiment, the DIMIC modulator comprises the nucleotide sequence of SEQ ID NOs 1-17 or 35-48, or a fragment thereof. In another embodiment, the DIMIC modulator is a DIM or a DIMIC polypeptide comprising the amino acid sequence of SEQ ID NOs 18-34 or 49-53, or a fragment thereof.

Examples of such DIMIC modulators are described earlier.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts an alignment of DIM1 protein sequences from various organisms. The amino acid residues differing from the consensus are shaded in a black box. **Plants:** At: *Arabidopsis thaliana*; Gm: *Glycine max*; Mt: *Medicago truncatula*; Le: *Lycopersicon esculentum*; Ga: *Gossypium arboreum*; Lj: *Lotus japonica*; Zm: *Zea mays*; Os: *Oryza sativa*; Pp: hybrid aspen (*Populus tremula* x *Populus tremuloides*); Pt: *Pinus taeda*; Hv: *Hordeum vulgare*; Ts: *Thellungiella salsuginea*; Cj: *Cryptomeria japonica*; Mc: *Mesembryanthemum crystallinum*; Ta: *Triticum aestivum*. **Fungi:** Sp: *Schizosaccharomyces pombe*. **Animals**: Dm: *Drosophila melanogaster*.

Figure 2 depicts the amino acid sequence of the DIMIC5 protein (SEQ ID NO 49). Indicated in the figure are the tandem WW/WWP domains (boxed, tryptophane and proline residues marked with an asterisk) separated by an 18-amino acid residue spacer. The C_2-domain is underscored by a rounded bracket. The DIMIC5 internal repeat domains are aligned as indicated by vertical lines connecting the conserved amino acid residues. The double underlined amino acid sequences represent the sequences of the GenBank entries with accession numbers AC004261 (protein ID AAD12009)/T02117. Amino acid residues not present in DIMIC5 are marked by a grey shaded box and are double underlined. The single underlined amino acid sequence corresponds to parts of the GenBank entry with accession number T02116. Extra amino acid residues present in T02116 but not in DIMIC5 or amino acid residues different between T02166 and DIMIC5 are indicated in a box. When amino acid sequence alignments are made identical residues are marked by a black box whereas similar residues (according to the groups (F,W,Y), (M,I,L,V), (R,K,H), (D,E), (N,Q), (S,T)) are marked by a grey shaded box. Gaps ('-') are introduced to obtain an optimal alignment.

Figure 3 depicts the genomic region of *Arabidopsis thaliana* (GenBank entry with accession number AC004261) comprising the DIMIC5 open reading frame (nucleotides 17241 to 20717 SEQ ID NO 37). Intron/exon positioning was modified (relative to the ORF predicted for the protein with ID AAD12009) to be in line with the experimentally determined partial DIMIC5 cDNA sequence (SEQ ID NO 35). Nucleotide residues marked by grey shaded boxes correspond to intron sequences. Bold-faced and underlined nucleotide residues correspond to the 5' extension added to complete the partial DIMIC5 cDNA. The 3' underlined nucleotide residues correspond to the 3' UTR of the partial DIMIC5 cDNA which are also part of AC004261 (nucleotides 20718 to 20924). The poly A^+ tail of the DIMIC5 cDNA is indicated between brackets purely for illustrative

reasons. Combination of the exon sequences yield the DIMIC5 ORF as partially present in the DIMIC5 cDNA (SEQ ID NO 36).

Figure 4 depicts the genomic region (SEQ ID NO 40) of *Arabidopsis thaliana* (GenBank entry with accession number AC008148) comprising the DIMIC7=DIMIC40 open reading frame (nucleotides 100439 to 106312: SEQ ID NO 39). Nucleotide residues marked by grey shaded boxes correspond to intron sequences. The 3' underlined nucleotide residues correspond to the partial DIMIC7 cDNA (SEQ ID NO 38) including the 3'UTR (the latter comprising nucleotides 106313 to 106564 which are also part of AC008148). The poly A^+ tail of the DIMIC7 cDNA is indicated between brackets purely for illustrative reasons. Combination of the exon sequences yield the DIMIC7=40 ORF as partially present in the DIMIC7=40 cDNA.

Figure 5 shows the amino acid sequence of the DIMIC7=DIMIC40 protein (SEQ ID NO 50) aligned with the *Arabidopsis thaliana* FAB1-like protein (AtFAB1; GenBank entry AL035525; protein ID CAB36798) and with homologous parts of the *Saccharomyces cerevisiae* (yeast) FAB1 kinase (GenBank entry P34756) as well as with the homologous part of the mouse CCTd protein (GenBank entry Z31554). Amino acid residues not present in mouse CCTd and not present in DIMIC7, AtFAB1 or yeast FAB1 are inserted were appropriate in block arrows. When amino acid sequence alignments are made identical residues are marked by a black box whereas similar residues (according to the groups (F,W,Y), (M,I,L,V), (R,K,H), (D,E), (N,Q), (S,T)) are marked by a grey shaded box. Gaps ('-') are introduced to obtain an optimal alignment. The amino acid residues corresponding to the partial DIMIC7=40 protein are underlined (SEQ ID NO 97). The activation loop of FAB1-type kinases is surrounded by a grey shaded box. Residues conserved in the C-terminal catalytic domain of Fab1-type kinases are marked with an asterisk. Among these are the invariant residues K2059, D2196 and D2216 (numbering relative to yeast FAB1) which are further marked by a surrounding box.

Figure 6 depicts the internal repeat domains found in the DIMIC7=DIMIC40 protein. Shown are the five different motifs (DIMIC7/1 to DIMIC7/5 corresponding with SEQ ID NOs 59 to 63) with indication of their position in the DIMIC7=40 protein sequence and the corresponding consensus sequence. When amino acid sequence alignments are made identical residues are marked by a black box whereas similar residues (according to the groups (F,W,Y), (M,I,L,V), (R,K,H), (D,E), (N,Q), (S,T)) are marked by a grey shaded box.

Figure 7 depicts the genomic region of *Arabidopsis thaliana* (GenBank entry with accession number AB023039) comprising the DIMIC26 open reading frame (nucleotides 19634 to 21435). Nucleotide residues marked by grey shaded boxes correspond to intron sequences. The 3' underlined nucleotide residues correspond to the partial DIMIC26 cDNA including the 3'UTR (the latter comprising nucleotides 19553 to 19633) which are also part of AB023039. The poly A^+ tail of the DIMIC26 cDNA is indicated between brackets purely for illustrative reasons. Combination of the exon sequences yield the DIMIC26 ORF as partially present in the DIMIC26 cDNA.

Figure 8 **(A)** depicts the amino acid sequence of the DIMIC26 protein (SEQ ID NO 51). The amino acid residues corresponding to the partial DIMIC26 protein (SEQ ID NO 106) are underlined. Figure 8 **(B)** depicts the '[M/I/L/V][R/K/H]' amino acid pair (double underlined) and the '[R/K/H][M/I/L/V]' amino acid pair (single underlined) which are repeated multiple times in the DIMIC26 protein. Note that both pairs can overlap.

Figure 9 depicts an alignment of the homologous amino acid regions of the DIMIC26 protein (SEQ ID NO 51) and the human centrosome protein E (CENP-E; GenBank accession number NM001813). When amino acid sequence alignments are made identical residues are marked by a black box whereas similar residues (according to the groups (F,W,Y), (M,I,L,V), (R,K,H), (D,E), (N,Q), (S,T)) are marked by a grey shaded box. Gaps ('-') are introduced to ensure optimal alignment. Further indicated in this figure are the '[M/I/L/V][R/K/H]' amino acid pair (double underlined) and the '[R/K/H][M/I/L/V]' amino acid pair (single underlined) which are repeated in both protein parts. Note that both pairs can overlap.

Figure 10 depicts an alignment of the homologous amino acid regions of the DIMIC26 protein (SEQ ID NO 51) and the human nonmuscle type B myosin heavy chain (NMMHC-B; GenBank accession number P35580). When amino acid sequence alignments are made identical residues are marked by a black box whereas similar residues (according to the groups (F,W,Y), (M,I,L,V), (R,K,H), (D,E), (N,Q), (S,T)) are marked by a grey shaded box. Gaps ('-') are introduced to ensure optimal alignment.

Figure 11 depicts the internal repeat domains found in the DIMIC26 protein. Shown are the six different motifs (DIMIC26/1 to DIMIC26/6 corresponding to SEQ ID NOs 64 to 69) with indication of their position in the DIMIC26 protein sequence and the corresponding consensus sequence. When amino acid sequence alignments are made identical residues are marked by a black box whereas similar residues (according to the groups (F,W,Y), (M,I,L,V), (R,K,H), (D,E), (N,Q), (S,T)) are marked by a grey shaded box.

Figure 12 represents the genomic region of *Arabidopsis thaliana* (GenBank entry with accession number AC007583) comprising the DIMIC70B and DIMIC70C open reading frame (nucleotides 64105 to 65587). Intron/exon positioning was modified (relative to the ORF predicted for the protein with ID AAF75085) to be in line with the experimentally determined partial DIMIC70B (SEQ ID NO 46) (and DIMIC70C (SEQ ID NO 94) cDNA sequence). **(A)** Nucleotide residues marked by grey shaded boxes correspond to intron sequences in respect of DIMIC70B. The 3' underlined nucleotide residues correspond to the partial DIMIC70B cDNA (SEQ ID NO 46) including the 3'UTR (the latter comprising nucleotides 63964 to 64104 which are also part of AC007583). The poly A^+ tail of the DIMIC70B cDNA is indicated between brackets purely for illustrative reasons. The bold-faced 'tga' nucleotide-triplet (nucleotides 64486-64484) is not present in the DIMIC70A cDNA (SEQ ID NO 45). Omission of the triplet thus results in the genomic sequence of the DIMIC70A allele (SEQ ID NO 47). **(B)** Nucleotide residues marked by grey shaded boxes correspond to intron sequences in respect of DIMIC70C.

Figure 13 shows the DIMIC70A cDNA sequence **(A)** (SEQ ID NO 44)as well as the DIMIC70A protein sequence **(B)** (SEQ ID NO 52). The underlined N-terminal extension added to complete the DIMIC70A protein is derived from GenBank entry AC007583/protein ID AAF75085. Amino acid residues not present in the protein with ID AAF75085 are bold-faced and indicated between brackets. (SEQ ID NO 108)

Figure 14 shows the DIMIC70B cDNA sequence (A) (SEQ ID NO 46) as well as the DIMIC70B protein sequence (B) (SEQ ID NO 53). The underlined N-terminal extension added to complete the DIMIC70B protein is derived from GenBank entry AC007583/protein ID AAF75085. Amino acid residues not present in said protein with ID AAF75085 are bold-faced and indicated between brackets (SEQ ID NO 109).

Figure 15 depicts an alignment of the C-terminal part of the DIMIC70A protein (SEQ ID NO 52) with the PRODOM family PD12637 consensus sequence which comprises the redox-active center of thioredoxins and thioredoxin-like proteins. The amino acid residues of DIMIC70A marked with an asterisk correspond to amino acid residues which are, according to the PD12637 consensus domain, preferably occurring at that given position. Amino acid residues of DIMIC70A marked with a '+' deviate from the preferably occurring amino acid residues. The N-terminal asterisks overlie the conserved thioredoxin-like 'CXXC' consensus motif. Further indicated are the redox-active centers of a number of proteins from different organisms including yeast TRG1 (Günther *et al.* (1991) *J Biol Chem* 266, 24557-24563), a number of protein disulfide isomerases (Wang

and Chang (1999) *EMBO J* 18, 5972-5982) and a number of thioredoxins and thioredoxin-like proteins as indicated with their GenBank accession and protein ID numbers. S The redox-active centers are aligned with the redox-active center of DIMIC70A and of the PRODOM PD12637 consensus sequence. When amino acid sequence alignments are made identical residues are marked by a black box whereas similar residues (according to the groups (F,W,Y), (M,I,L,V), (R,K,H), (D,E), (N,Q), (S,T)) are marked by a grey shaded box.

Figure 16 shows the DIMIC70C cDNA sequence **(A)** (SEQ ID NO 94) as well as the DIMIC70C protein sequence **(B)** (SEQ ID NO 95). The start and stop codons are represented in bold in the nucleotide sequence. The first 17 amino acids of the deduced protein sequence indicated in bold differ from the DIMIC70B amino acid sequence.

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as "DIM1-interacting molecules" or "DIMIC" nucleic acid and polypeptide molecules. The DIMIC molecules of the present invention were identified based on their ability, as determined using yeast two-hybrid assays (described in detail in Example 2), to interact with the protein AtDIM1, the DIM1 homolog from *Arabidopsis thaliana* as well as with other DIM1 homologs such as those of other plants (described in detail in Example 1).

DIM1 is involved in one or more of the following processes: (a) Cell cycle processes including, but not limited to, processes associated with G2/M transition or chromosome movement and segregation, spindle formation and elongation, cytokinesis, and regulation of the APC/C (Berry and Gould (1997) *J Cell Biol* 137, 1337-1354; Berry *et al.* (1999), *Mol Cell Biol* 19, 2535-2546; Law *et al.* (1998) *Mol Cell Biol* 18, 3540-3551; and see Example 5); (b) Pre-mRNA splicing (Stevens and Abelson (1999) *Proc Natl Acad Sci USA* 96, 7226-7231; Teigelkamp *et al.* (1998) *RNA* 4, 127-141; Zhang *et al.* (1999) *Physiol Genomics* 1, 109-118); (c) Vesicle transport or processing (see Examples 3 and 4).

Because of their ability to interact with (*e.g.*, bind to) AtDIM1 and possibly AtDIM1 homologues (see Example 1), the DIMIC molecules of the present invention may modulate, *e.g.*, upregulate or downregulate, the activity of DIM1. Furthermore, because of their ability to interact with (*e.g.*, bind to) AtDIM1 and possibly AtDIM1 homologues which are proteins involved in cell cycle regulation and/or pre-mRNA splicing and/or vesicle transport/processing, the DIMIC molecules of the present invention may also play a role in cell cycle regulation and/or pre-mRNA splicing and/or vesicle transport/processing in, for example, plant or animal cells.

As used herein, the terms "DIM1-interacting protein" or "DIMIC" include a polypeptide which interacts with, *e.g.*, binds to a DIM1 protein, and which is involved in controlling or regulating the cell cycle and/or pre-mRNA splicing and/or vesicle transport/processing, or part of any of these processes, in a cell, tissue, organ or in a whole organism. DIMIC molecules of the present invention may also be capable of binding to, regulating, or being regulated by cyclin-dependent kinases, such as plant cyclin dependent kinases, *e.g.*, CDC2a or CDC2b, or their subunits. The term DIMIC also includes fragments, variants, homologs, alleles or precursors (*e.g.*, pre-proteins, pre-pro-proteins or pro-proteins) of DIMIC polypeptides.

As used herein, the term "cell cycle" includes the cyclic biochemical and structural events associated with the growth, division and proliferation of cells, and in particular with the regulation of the replication of DNA and mitosis. The cell cycle is divided into periods or phases called: G0, Gap1 (G1), DNA synthesis (S), Gap2 (G2), and mitosis (M). Normally these four phases occur sequentially, however, the term "cell cycle" also includes modified cycles wherein one or more phases are absent resulting in modified cell cycle such as endomitosis, acytokinesis, polyploidy, polyteny, and endoreduplication.

As used herein, the term "pre-mRNA splicing" includes the biochemical events associated with the nuclear processing of eukaryotic pre-mRNA leading to their conversion into mature mRNA species competent for translation into a protein. "Pre-mRNA splicing" is effectuated by small ribonucleoprotein (snRNPs) particles in association with several non-snRNPs (Staley and Guthrie (1998) *Cell* 92, 315-326), including DIM1 (Stevens and Abelson (1999) *Proc Natl Acad Sci USA* 96, 7226-7231).

As used herein, the term "vesicle transport/processing" includes all processes leading to the formation, transport, processing and fusion of cellular organelles surrounded by a phospholipid membrane as well as to whole cells surrounded by a phospholipid membrane. The term"vesicle transport/processing" further includes the biosynthesis, transport, processing and degradation of components of the phospholipid membranes as well as of non-phospholipid components, *e.g.*, membrane proteins or lipoproteins, carried within said phospholipid membranes or within the lumen of vesicles surrounded by said membranes.

As used herein, the term "plant" includes whole plants, plant organs (*e.g.*, leaves, stems, or roots), plant tissue, plant seeds, and plant cells and progeny thereof. The class of plants which can be used in the methods of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants. Particularly preferred plants are *Arabidopsis thaliana*, rice, wheat, barley, sorghum, maize, tomato, potato, cotton, alfalfa, oilseed rape, soybean, cotton, sunflower or canola. The term plant also includes monocotyledonous (monocot) plants and dicotyledonous (dicot) plants including a fodder or forage legume, ornamental plants, food crops, trees, or shrubs selected from the list comprising *Acacia spp., Acer spp., Actinidia spp., Aesculus spp., Agathis australis, Albizia amara, Alsophila tricolor, Andropogon spp., Arachis spp, Areca catechu, Astelia fragrans, Astragalus cicer, Baikiaea plurijuga, Betula spp., Brassica spp., Bruguiera gymnorrhiza, Burkea africana, Butea frondosa, Cadaba farinosa, Calliandra spp, Camellia sinensis, Canna indica, Capsicum spp., Cassia spp., Centroema pubescens, Chaenomeles*

spp.,Cinnamomum cassia, Coffea arabica, Colophospermum mopane, Coronillia varia, Cotoneaster serotina, Crataegus spp., Cucumis spp., Cupressus spp., Cyathea dealbata, Cydonia oblonga, Cryptomeria japonica, Cymbopogon spp., Cynthea dealbata, Cydonia oblonga, Dalbergia monetaria, Davallia divaricata, Desmodium spp., Dicksonia squarosa, Diheteropogon amplectens, Dioclea spp, Dolichos spp., Dorycnium rectum, Echinochloa pyramidalis, Ehrartia spp., Eleusine coracana, Eragrestis spp., Erythrina spp., Eucalyptus spp., Euclea schimperi, Eulalia villosa, Fagopyrum spp., Feijoa sellowiana, Fragaria spp., Flemingia spp, Freycinetia banksii, Geranium thunbergii, Ginkgo biloba, Glycine javanica, Gliricidia spp, Gossypium hirsutum, Grevillea spp., Guibourtia coleosperma, Hedysarum spp., Hemarthia altissima, Heteropogon contortus, Hordeum vulgare, Hyparrhenia rufa, Hypericum erectum, Hyperthelia dissoluta, Indigo incarnata, Iris spp., Leptarrhena pyrolifolia, Lespediza spp., Lettuca spp., Leucaena leucocephala, Loudetia simplex, Lotonus bainesii, Lotus spp., Macrotyloma axillare, Malus spp., Manihot esculenta, Medicago sativa, Metasequoia glyptostroboides, Musa sapientum, Nicotianum spp., Onobrychis spp., Ornithopus spp., Oryza spp., Peltophorum africanum, Pennisetum spp., Persea gratissima, Petunia spp., Phaseolus spp., Phoenix canariensis, Phormium cookianum, Photinia spp., Picea glauca, Pinus spp., Pisum sativum, Podocarpus totara, Pogonarthria fleckii, Pogonarthria squarrosa, Populus spp., Prosopis cineraria, Pseudotsuga menziesii, Pterolobium stellatum, Pyrus communis, Quercus spp., Rhaphiolepsis umbellata, Rhopalostylis sapida, Rhus natalensis, Ribes grossularia, Ribes spp., Robinia pseudoacacia, Rosa spp., Rubus spp., Salix spp., Schyzachyrium sanguineum, Sciadopitys verticillata, Sequoia sempervirens, Sequoiadendron giganteum, Sorghum bicolor, Spinacia spp., Sporobolus fimbriatus, Stiburus alopecuroides, Stylosanthos humilis, Tadehagi spp, Taxodium distichum, Themeda triandra, Trifolium spp., Triticum spp., Tsuga heterophylla, Vaccinium spp., Vicia spp., Vitis vinifera, Watsonia pyramidata, Zantedeschia aethiopica, Zea mays, amaranth, artichoke, asparagus, broccoli, brussel sprout, cabbage, canola, carrot, cauliflower, celery, collard greens, flax, kale, lentil, oilseed rape, okra, onion, potato, rice, soybean, straw, sugarbeet, sugar cane, sunflower, tomato, squash, and tea, amongst others, or the seeds of any plant specifically named above or a tissue, cell or organ culture of any of the above species.

The term "plant cell", as used herein includes seeds, seed suspension cultures, embryos, cells from meristematic regions, cells from callus tissue, cells from leaves, cells from roots, cells from shoots, gametophytes, sporophytes, pollen, and microspores

The DIMIC molecules of the present invention are involved in the regulation of cell cycle and/or pre-mRNA splicing and/or vesicle transport/processing, or part of any of these processes in plants, fungi and animals. Accordingly, the DIMIC molecules of the present invention, or derivatives thereof, may be used to modulate the cell cycle and/or pre-mRNA splicing and/or vesicle transport/processing, or part of any of these processes in an organism by, for example, modulating the activity or level of expression of a DIMIC molecule of the present invention. In plants, the DIMIC molecules of the present invention may be used in agriculture to, for example, improve the growth characteristics of a plant such as the growth rate of a plant; the size of specific tissues or organs in a plant; or the architecture or morphology of a plant. The DIMIC molecules of the present invention may also be used in agriculture to increase crop yield, improve tolerance to environmental stress conditions (such as drought, salt, temperature, or nutrient deprivation), improve tolerance to plant pathogens that abuse the cell cycle, or as targets to facilitate the identification of inhibitors or activators of DIMs or DIMICs that may be useful as phytopharmaceuticals, herbicides or plant growth regulators. The DIMIC molecules of the present invention may also be used, *e.g.*, in agriculture, to treat a cell cycle associated disorder.

As used herein, the term "cell cycle associated disorder" includes a disorder, disease or condition which is caused or characterized by a misregulation (*e.g.*, downregulation or upregulation), abuse, arrest, or modification of the cell cycle. In plants cell cycle associated disorders include endomitosis, acytokinesis, polyploidy, polyteny, and endoreduplication which may be caused by external factors such as pathogens (nematodes, viruses, fungi, or insects), chemicals, environmental stress (*e.g.*, drought, temperature, nutrients, or UV light) resulting in, for example, neoplastic tissue (*e.g.*, galls, root knots) or inhibition of cell division/proliferation (*e.g.*, stunted growth). Cell cycle associated disorders in animals include proliferative disorders or differentiative disorders, such as cancer, *e.g.*, melanoma, prostate cancer, servical cancer, breast cancer, colon cancer, or sarcoma.

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as DIMIC protein and nucleic acid molecules, which comprise a family of molecules having certain conserved structural and functional features. The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-

naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of plant, *e.g.* Arabidopsis, origin, as well as other, distinct proteins of plant, *e.g.*, Arabidopsis, origin or alternatively, can contain homologues of other plants, *e.g.*, rice, or of non-plant origin. Members of a family may also have common functional characteristics.

In one embodiment of the invention, a DIMIC protein of the present invention, is identified based on the presence of at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fiveteen, at least sixteen, at least seventeen, at least eightteen, at least nineteen, at least twenty, or at least more more of the following motifs:

A. FKBP domain

As used herein, the term "FKBP domain" includes a domain of about 16-20 amino acid residues in length and which has the following consensus pattern: [L/I/V/M/C] X [Y/F] X [G/V/L] X_{1-2} [L/F/T] X_2 G X_3 [D/E] [S/T/A/E/Q/K] [S/T/A/N] (SEQ ID NO 54), with 'X' being any amino acid residue, "X_n" being a stretch of "n" random amino acid residues and, *e.g.*, '[Y/F]' meaning either a tyrosine or phenylalanine residue occurring at that position.

B. WW or WWP domain

As used herein, the term "WW domain" or "WWP domain" includes a domain of about 27-30 amino acid residues in length and which has the following consensus pattern:

$WX_{22}WX_2P$ (SEQ ID NO 55), with "X" being any amino acid and "X_n" being a stretch of n of Xs. WW domains are, typically, small and compact globular structures that interact with proline-rich ligands (Bedford *et al.* (1997) *EMBO J* 16, 2376-2383; Chan *et al.* (1996) *EMBO J* 15, 1045-1054; Einbond and Sudol (1996) *FEBS Lett* 384, 1-8).

C. Non-classical C_2-domain

As used herein, the term "non-classical C_2-domain" includes a C_2-domain as present in human and mouse polyglutamine tract-binding protein (PQBP-1) and includes a domain of about 30-35, preferably about 32-33, amino acid residues in length and which has the following consensus pattern:

KKX_5D[D/E]ELDPMDPSSYSDAPRGX$WX_2GLX_{0-1}K$ (SEQ ID NO 56)

with X being any amino acid and Xn being a stretch of n of Xs and [D/E] being either an aspartate or glutamate residue at that position.

Most proteins containing C_2-domains are functional in signal transduction or membrane traffic. Phospholipid binding to many C_2-domains is regulated by Ca^{2+} and, therefore, C_2-domain proteins are implicated in Ca^{2+}-dependent phospholipid signalling (Rizo and Südhof (1998) *J Biol Chem* 273, 15879-15882).

D. DIMIC5 internal repeat domain

As used herein, the term "DIMIC5 internal repeat domain" includes a domain of about 5-10, preferably 7, amino acid residues in length and which has the following consensus pattern: GGWXVGL (SEQ ID NO 57) with X being any amino acid.

E. FAB1 activation loop

As used herein, the term "FAB1 activation loop" includes a domain of about 18-22, preferably 19, amino acid residues in length and which has the following consensus pattern: T[F/Y]T[W/L]DKKLE[S/T/M]WVKXXG[I/L][V/L]G (SEQ ID NO 58)
with the with X being any amino acid.

This motif may be involved in defining PtdIns3*P* as the substrate for 5-phosphorylation (McEwen *et al.* (1999) *J. Biol. Chem.* 274, 33905-33912).

F. DIMIC7 internal repeat domains

As used herein, the term "DIMIC7 internal repeat domain" or "motif DIMIC7/N" includes domains (numbered by 'N' in DIMIC7/N annotation) of about 6-8 amino acid residues in length and which have one of the following consensus patterns:
Motif DIMIC7/1: PLGR[F/W/Y][M/I/L/V] (SEQ ID NO 59);
Motif DIMIC7/2: EXXG[R/K/H]IW (SEQ ID NO 60);
Motif DIMIC7/3: DLXXPT[M/I/L/V] (SEQ ID NO 61);
Motif DIMIC7/4: DDXXSXYF (SEQ ID NO 62);and
Motif DIMIC7/5: TEXSDXLN (SEQ ID NO 63); with X being any amino acid and, *e.g.*, [D/E] being either an aspartate or glutamate residue at that position.

G. DIMIC26 internal repeat domains

As used herein, the term "DIMIC26 internal repeat domain" or "motif DIMIC26/N" includes domains (numbered by 'N' in DIMIC26/N annotation) of about 6-9 amino acid residues in length and which have one of the following consensus patterns:

Motif DIMIC26/1: CXCXIC (SEQ ID NO 64);
Motif DIMIC26/2: ACNRXXE[M/I/L/V][M/I/L/V](SEQ ID NO 65);
Motif DIMIC26/3: QXSGGG (SEQ ID NO 66);
Motif DIMIC26/4: [M/I/L/V]DX[M/I/L/V]KXGL (SEQ ID NO 67);
Motif DIMIC26/5: SEXXAEKQ(SEQ ID NO 68); and
Motif DIMIC26/6: RLXXAEA[D/E](SEQ ID NO 69); with X being any amino acid and, *e.g.*, [D/E] being either an aspartate or glutamate residue at that position.

H. DIMIC26 di-amino acid motifs

As used herein, the term "DIMIC26 di-amino acid motifs" includes domains of 2 amino acid residues in length and which have the following consensus patterns: [M/I/L/V] [R/K/H]; and [R/K/H] [M/I/L/V] with, *e.g.*, [R/K/H] being either an arginine, lysine or histidine residue at that position.

I. Thioredoxin-like domain

As used herein, the term "thioredoxin-like domain" includes a domain of about 4 amino acid residues in length and which has the following consensus patterns:
CXXC (SEQ ID NO 70) (Wang and Chang (1999) *EMBO J* 18, 5972-5982)
with X being any amino acid.

J. PEST sequence

As used herein, the term "PEST sequence" includes an amino acid domain of variable length which is enriched in the amino acid residues proline, glutamate, serine and/or threonine. Potential PEST sequences can be identified using the PESTFIND software (can be downloaded from http://www.ebi.ac.uk). The presence of a PEST sequence in a protein is indicative of a high turnover rate, *i.e.*, low stability or short half-life, of said protein (Rogers *et al* (1986) *Science* 234, 364-368).

K. PHD finger

The PHD finger is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation (Gibson *et al* (1995) *Trends Biochem. Sci.* 20: 56-59). The PHD finger motif is reminiscent of, but distinct from the C3HC4 type RING finger. The function of this domain is not yet known but in analogy with the LIM domain it could be involved in protein-protein interaction and be important for

the assembly or activity of multicomponent complexes involved in transcriptional activation or repression. In similarity to the RING finger and the LIM domain, the PHD finger is thought to bind two zinc ions.

Isolated DIMIC proteins of the present invention have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NOs 49-53 and 95 or are encoded by a nucleotide sequence sufficiently identical to SEQ ID NOs 35-48 or 94.

As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least 30%, 40%, or 50% identity, preferably 60% identity, more preferably 70%-80%, and even more preferably 90-95% identity across the amino acid sequences of the domains and contain at least one, or at least two or three or four structural domains or motifs, are defined herein as sufficiently identical. Furthermore, amino acid or nucleotide sequences which share at least 30%, 40%, or 50%, preferably 60%, more preferably 70-80%, or 90-95% identity and share a common functional activity are defined herein as sufficiently identical.

As used interchangeably herein, a "DIMIC activity", "biological activity of DIMIC", or "functional activity of DIMIC", refers to an activity exerted by a DIMIC protein, polypeptide or nucleic acid molecule on a DIMIC responsive cell or tissue, or on a DIMIC protein substrate, as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, a DIMIC activity is a direct activity, such as an association with DIMIC-target molecule, *e.g.*, DIM1. As used herein, a "target molecule" or "binding partner" is a molecule with which a DIMIC protein binds or interacts in nature, such that DIMIC-mediated function is achieved. A DIMIC target molecule can be a non-DIMIC molecule, or a DIMIC protein or polypeptide of the present invention. In an exemplary embodiment, a DIMIC target molecule is a DIMIC ligand. Alternatively, a DIMIC activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the DIMIC protein with a DIMIC ligand. The biological activities of DIMICs are described herein.

For example, the DIMIC proteins of the present invention can have one or more of the following functions: (1) they may act in the cell cycle, more specifically in cell cycle processes including but not limited to G2/M transition or chromosome movement and

segregation, spindle formation and elongation, cytokinesis, or regulation of the APC/C; (2) they may modulate pre-mRNA splicing; and (3) they may modulate vesicle transport or processing.

Accordingly, another embodiment of the present invention features isolated DIMIC proteins and polypeptides having a DIMIC activity. Preferred proteins are DIMIC proteins, *e.g.*, DIMIC proteins from plants, having at least one or more of the following domains: a "WW or WWP domain", a "non-classical C_2-domain", a "FAB1 activation loop", a "DIMIC5 internal repeat domain", a "DIMIC7 internal repeat domain", a "DIMIC26 internal repeat domain", a "DIMIC26 di-amino acid motif", a "thioredoxin-like domain" and/or a "PEST sequence", and, preferably, a DIMIC activity activity.

Additional preferred proteins, *e.g.*, DIMIC proteins from plants, have at least one or more of the following domains: a "WW or WWP domain", a "non-classical C_2-domain", a "FAB1 activation loop", a "DIMIC5 internal repeat domain", a "DIMIC7 internal repeat domain", a "DIMIC26 internal repeat domain", a "DIMIC26 di-amino acid motif", a "thioredoxin-like domain" and/or a "PEST sequence" and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOs 35-48.

The sequences of the present invention are summarized below, in Table I.

TABLE I

DIMIC#	SEQ ID NO: partial cDNA	SEQ ID NO: full-length cDNA	SEQ ID NO: genomic DNA	SEQ ID NO: full-length protein
DIMIC5	35	36	37	49
DIMIC7=40	38	39	40	50
DIMIC26	41	42	43	51
DIMIC70A	44	45	47	52
DIMIC70B		46	48	53
DIMIC70C		94		95

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode DIMIC proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify DIMIC-encoding nucleic acids (*e.g.*, DIMIC mRNA) and fragments for use as PCR primers for the amplification or mutation of DIMIC nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated DIMIC nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NOs 35-48 or 94, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, using all or portion of the nucleic acid sequence of SEQ ID NOs 35-48 or 94, as a hybridization probe, DIMIC nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold*

Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NOS 35-48 or 94can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NOS 35-48, respectively.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be

cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to DIMIC nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NOs 35-48 or 94.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NOs35-48, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NOs 35-48 or 94, is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NOs 35-48 or 94, respectively, such that it can hybridize to the nucleotide sequence shown in SEQ ID NOs 35-48 or 94, respectively, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to the nucleotide sequence (*e.g.*, to the entire length of the nucleotide sequence) shown in SEQ ID NOs 35-48 or 94, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NOs 35-48 or 94, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a DIMIC protein. The nucleotide sequence determined from the cloning of the DIMIC gene allows for the generation of probes and primers designed for use in identifying and/or cloning other DIMIC family members, as well as DIMIC homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The

oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense sequence of SEQ ID NOs 35-48 or 94, or of a naturally occurring allelic variant or mutant of SEQ ID NOs 35-48. In an exemplary embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, or 800 nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NOs 35-48 or 94.

Probes based on the DIMIC nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which misexpress a DIMIC protein, such as by measuring a level of a DIMIC-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting DIMIC mRNA levels or determining whether a genomic DIMIC gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of a DIMIC protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NOs 35-48 or 94, which encodes a polypeptide having a DIMIC biological activity (the biological activities of the DIMIC proteins are described herein), expressing the encoded portion of the DIMIC protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the DIMIC protein.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NOs 35-48 or 94, due to the degeneracy of the genetic code and, thus, encode the same DIMIC proteins as those encoded by the nucleotide sequence shown in SEQ ID NOs 35-48 or 94. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a DIMIC protein.

In addition to the DIMIC nucleotide sequences shown in SEQ ID NOs 35-48 or 94, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the DIMIC proteins may exist within a population (*e.g.*, an *Arabidopsis* or rice plant population). Such genetic polymorphism in the DIMIC genes may exist among individuals within a population due to natural allelic

variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding an DIMIC protein, preferably a plant DIMIC protein, and can further include non-coding regulatory sequences, and introns. Such natural allelic variations include both functional and non-functional DIMIC proteins and can typically result in 1-5% variance in the nucleotide sequence of a DIMIC gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in DIMIC genes that are the result of natural allelic variation and that do not alter the functional activity of a DIMIC protein are intended to be within the scope of the invention.

Natural allelic variants are further include molecules that comprise single nucleotide polymorphisms (SNPs) as well as small insertion/deletion polymorphisms (INDELs; the size of INDELs is usually less than about 100 bp). SNPs and INDELs form the largest set of sequence variants in naturally occurring polymorphic strains of most organisms. They are helpful in mapping genes and in discovery of genes and gene functions. They are furthermore helpful in the identification of genetic loci, *e.g.*, plant genes, involved in determining processes such as growth rate, plant size and plant yield, plant vigor, disease resistance, stress tolerance and the like. Many techniques are nowadays available to identify SNPs and/or INDELs including (i) PCR followed by denaturing high performance liquid chromatography (DHPLC; *e.g.*, Cho *et al.* (1999) *Nature Genet* 23, 203-207); (ii) constant denaturant capillary electrophoresis (CDCE) combined with high-fidelity PCR (*e.g.*, Li-Sucholeiki *et al.* (1999) *Electrophoresis* 20, 1224-1232); (iii) denaturing gradient gel electrophoresis (Fischer and Lerman (1983) *Proc. Natl. Acad. Sci. USA* 80, 1579-1583); (iv) matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS; *e.g.*, Ross *et al* (2000) *Biotechniques* 29, 620-629); (v) real-time fluorescence monitoring PCR assays (Tapp *et al* (2000) *Biotechniques* 28, 732-738); (vi) AcryditeTM gel technology (Kenney *et al* (1998) *Biotechniques* 25, 516-521); (vii) cycle dideoxy fingerprinting (CddF; Langemeier *et al* (1994) *Biotechniques* 17, 484-490); (viii) single-strand conformation polymorphism (SSCP) analysis (Vidal-Puig and Moller (1994) *Biotechniques* 17, 490-496) and (ix) mini-sequencing primer extension reaction (Syvanen (1999) *Hum Mutat* 13, 1-10). The technique of 'Targeting Induced Local Lesions in Genomes' (TILLING; McCallum *et al.* (2000) *Nat. Biotechnol* 18, 455-457; *Plant Physiol* 123, 439-442), which is a variant of (i) *supra*, can also be applied to rapidly identify an altered gene in, *e.g.*, chemically mutagenized plant individuals showing interesting phenotypes.

Differences in preferred codon usage are illustrated below for *Agrobacterium tumefaciens* (a bacterium), *Arabidopsis thaliana*, *Medicago sativa* (two dicotyledonous

plants) and *Oryza sativa* (a monocotyledonous plant). These examples were extracted from http://www.kazusa.or.jp/codon. For example, the codon GGC (for glycine) is the most frequently used codon in *A. tumefaciens* (36.2 ‰), is the second most frequently used codon in *O. sativa* but is used at much lower frequencies in *A. thaliana* and *M. sativa* (9 ‰ and 8.4 ‰ , respectively). Of the four possible codons encoding glycine the GGC codon is most preferably used in *A. tumefaciens* and *O. sativa.* However, in *A. thaliana* the GGA (and GGU) codon is most preferably used, whereas in *M. sativa* the GGU (and GGA) codon is most preferably used.

Moreover, nucleic acid molecules encoding other DIMIC family members and, thus, which have a nucleotide sequence which differs from the DIMIC sequences of SEQ ID NOs 35-48 or 94 are intended to be within the scope of the invention. For example, another DIMIC cDNA can be identified based on the nucleotide sequence of the plant DIMIC moleculels described herein. Moreover, nucleic acid molecules encoding DIMIC proteins from different species, and thus which have a nucleotide sequence which differs from the DIMIC sequences of SEQ ID NOs 35-48 or 94 are intended to be within the scope of the invention. For example, a human DIMIC cDNA can be identified based on the nucleotide sequence of a plant DIMIC.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the DIMIC cDNAs of the invention can be isolated based on their homology to the DIMIC nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOs 35-48. In other embodiment, the nucleic acid is at least 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, or 600 nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 30%, 40%, 50%, or 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization

conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50°C, preferably at 55°C, more preferably at 60°C, and even more preferably at 65°C. Ranges intermediate to the above-recited values, *e.g.*, at 60-65 °C or at 55-60 °C are also intended to be encompassed by the present invention. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NOs 35-48 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In addition to naturally-occurring allelic variants of the DIMIC sequences that may exist in nature, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NOs 35-48 or 94, thereby leading to changes in the amino acid sequence of the encoded DIMIC proteins, without altering the functional ability of the DIMIC proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of a DIMIC protein. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of DIMIC without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the DIMIC proteins of the present invention, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the DIMIC proteins of the present invention and other DIMIC family members are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding DIMIC proteins that contain changes in amino acid residues that are not essential for activity.

An isolated nucleic acid molecule encoding a DIMIC protein homologous to the DIMIC proteins of the present invention can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOs 35-48 or 94, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NOs 35-48 or 94 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at

one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a DIMIC protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a DIMIC coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for DIMIC biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NOs 35-48 or 94, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant DIMIC protein can be assayed for the ability to: (1) modulate cell cycle processes including but not limited to G2/M transition or chromosome movement and segregation, spindle formation and elongation, cytokinesis, or regulation of the APC/C; (2) modulate pre-mRNA splicing; (3) modulate vesicle transport or processing; or (4) interact with DIM1 in, *e.g.*, a yeast two hybrid assay.

In addition to the nucleic acid molecules encoding DIMIC proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire DIMIC coding strand, or only to a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding DIMIC. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence

encoding DIMIC. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding DIMIC disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of DIMIC mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of DIMIC mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of DIMIC mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

Preferably, production of antisense nucleic acids in plants occurs by means of a stably integrated transgene comprising a promoter operative in plants, an antisense oligonucleotide, and a terminator. Other known nucleotide modifications include methylation, cyclization and 'caps' and substitution of one or more of the naturally occurring nucleotides with an analog such as inosine. Modifications of nucleotides include the addition of acridine, amine, biotin, cascade blue, cholesterol, Cy3®, Cy5®, Cy5.5® Dabcyl, digoxigenin, dinitrophenyl, Edans, 6-FAM, fluorescein, 3'-glyceryl, HEX, IRD-700, IRD-800, JOE, phosphate psoralen, rhodamine, ROX, thiol (SH), spacers, TAMRA, TET, AMCA-S®, SE, BODIPY®, Marina Blue®, Pacific Blue®, Oregon Green®, Rhodamine Green®, Rhodamine Red®, Rhodol Green® and Texas Red®. Polynucleotide backbone modifications include methylphosphonate, 2'-OMe-methylphosphonate RNA, phosphorothiorate, RNA, 2'-OMeRNA. Base modifications include 2-amino-dA, 2-aminopurine, 3'-(ddA), 3'dA(cordycepin), 7-deaza-dA, 8-Br-dA, 8-oxo-dA, N^6-Me-dA, abasic site (dSpacer), biotin dT, 2'-OMe-5Me-C, 2'-OMe-propynyl-C, 3'-(5-Me-dC), 3'-(ddC), 5-Br-dC, 5-I-dC, 5-Me-dC, 5-F-dC, carboxy-dT, convertible dA, convertible dC, convertible dG, convertible dT, convertible dU, 7-deaza-dG, 8-Br-dG, 8-oxo-dG, O^6-Me-dG, S6-DNP-dG, 4-methyl-indole, 5-nitroindole, 2'-OMe-inosine, 2'-dI, 0^6-phenyl-dI, 4-methyl-indole, 2'-deoxynebularine, 5-nitroindole, 2-aminopurine, dP(purine analogue), dK(pyrimidine analogue), 3-nitropyrrole, 2-thio-dT, 4-thio-dT, biotin-dT, carboxy-dT, O^4-Me-dT, O^4-triazol dT, 2'-OMe-propynyl-U, 5-Br-dU, 2'-dU, 5-F-dU, 5-I-dU, O^4-triazol dU.

The antisense nucleic acid molecules of the invention are typically introduced into a plant or administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a DIMIC protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of introduction or administration of antisense nucleic acid molecules of the invention include transformation in a plant or direct injection at a tissue site in a subject. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the

vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a constitutive promoter or a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In another embodiment, the antisense nucleic acid molecule further comprises a sense nucleic acid molecule complementary to the antisense nucleic acid molecule. Gene silencing methods based on such nucleic acid molecules are well known to the skilled artisan (*e.g.*, Grierson *et al.* (1998) WO 98/53083; Waterhouse *et al.* (1999) WO 99/53050).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave DIMIC mRNA transcripts to thereby inhibit translation of DIMIC mRNA. A ribozyme having specificity for a DIMIC-encoding nucleic acid can be designed based upon the nucleotide sequence of a DIMIC cDNA disclosed herein (*i.e.*, SEQ ID NOs 35-48). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a DIMIC-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, DIMIC mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

The use of ribozymes for gene silencing in plants is known in the art (*e.g.*, Atkins *et al.* (1994) WO 94/00012; Lenne *et al.* (1995) WO 95/03404; Lutziger *et al.* (2000) WO 00/00619; Prinsen *et al.* (1997) WO 97/13865 and Scott *et al.* (1997) WO/ 97/38116).

Alternatively, DIMIC gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the DIMIC (*e.g.*, the DIMIC promoter and/or enhancers) to form triple helical structures that prevent transcription of the DIMIC gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the DIMIC nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* Proc. Natl. Acad. Sci. 93: 14670-675.

PNAs of DIMIC nucleic acid molecules can be used for increasing crop yield in plants or in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of DIMIC nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of DIMIC can be modified, (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of DIMIC nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (*e.g.*, RNAse H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high

binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. US.* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (see, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

II. Isolated DIMIC Proteins and Anti-DIMIC Antibodies

One aspect of the invention pertains to isolated DIMIC proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-DIMIC antibodies. In one embodiment, native DIMIC proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, DIMIC proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a DIMIC protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the DIMIC protein is derived, or substantially free from chemical

precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of DIMIC protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of DIMIC protein having less than about 30% (by dry weight) of non-DIMIC protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-DIMIC protein, still more preferably less than about 10% of non-DIMIC protein, and most preferably less than about 5% non-DIMIC protein. When the DIMIC protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of DIMIC protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of DIMIC protein having less than about 30% (by dry weight) of chemical precursors or non-DIMIC chemicals, more preferably less than about 20% chemical precursors or non-DIMIC chemicals, still more preferably less than about 10% chemical precursors or non-DIMIC chemicals, and most preferably less than about 5% chemical precursors or non-DIMIC chemicals.

Biologically active portions of a DIMIC protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the DIMIC protein, which include less amino acids than the full length DIMIC proteins, and exhibit at least one activity of a DIMIC protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the DIMIC protein. A biologically active portion of a DIMIC protein can be a polypeptide which is, for example, at least 10, 25, 50, 100 or more amino acids in length.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%,

80%, or 90% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to DIMIC nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to DIMIC protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The invention also provides DIMIC chimeric or fusion proteins. As used herein, a DIMIC "chimeric protein" or "fusion protein" comprises a DIMIC polypeptide operatively linked to a non-DIMIC polypeptide. An "DIMIC polypeptide" refers to a polypeptide having

an amino acid sequence corresponding to DIMIC, whereas a "non-DIMIC polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the DIMIC protein, *e.g.*, a protein which is different from the DIMIC protein and which is derived from the same or a different organism. The non- DIMIC polypeptide can, for example, be $(histidine)_6$-tag, glutathione S-transferase, protein A, maltose-binding protein, dihydrofolate reductase, Tag•100 epitope (EETARFQPGYRS; SEQ ID NO 75), c-myc epitope (EQKLISEEDL; SEQ ID NO 76), FLAG®-epitope (DYKDDDK; SEQ ID NO 77), lacZ, CMP (calmodulin-binding peptide), HA epitope (YPYDVPDYA; SEQ ID NO 78), protein C epitope (EDQVDPRLIDGK; SEQ ID NO 79) or VSV epitope (YTDIEMNRLGK; SEQ ID NO 80).

Within a DIMIC fusion protein the DIMIC polypeptide can correspond to all or a portion of a DIMIC protein. In a preferred embodiment, a DIMIC fusion protein comprises at least one biologically active portion of a DIMIC protein. In another preferred embodiment, a DIMIC fusion protein comprises at least two biologically active portions of a DIMIC protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the DIMIC polypeptide and the non-DIMIC polypeptide are fused in-frame to each other. The non-DIMIC polypeptide can be fused to the N-terminus or C-terminus of the DIMIC polypeptide.

For example, in one embodiment, the fusion protein is a GST-DIMIC fusion protein in which the DIMIC sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant DIMIC.

In another embodiment, the fusion protein is a DIMIC protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, plant or mammalian host cells), expression and/or secretion of DIMIC can be increased through use of a heterologous signal sequence.

The DIMIC fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a plant or a subject *in vivo*. The DIMIC fusion proteins can be used to affect the bioavailability of a DIMIC substrate. Use of DIMIC fusion proteins may be useful agriculturally for the increase of crop yields or therapeutically for the treatment of cellular growth related disorders, *e.g.*, cancer. Moreover, the DIMIC-fusion proteins of the invention can be used as immunogens to produce anti-DIMIC antibodies in a subject, to purify DIMIC ligands and in screening assays to identify molecules which inhibit the interaction of DIMIC with a DIMIC substrate, *e.g.*, a kinase such as CDC2b.

Preferably, a DIMIC chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A DIMIC-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the DIMIC protein.

The present invention also pertains to variants of the DIMIC proteins which function as either DIMIC agonists (mimetics) or as DIMIC antagonists. Variants of the DIMIC proteins can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of a DIMIC protein. An agonist of the DIMIC proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a DIMIC protein. An antagonist of a DIMIC protein can inhibit one or more of the activities of the naturally occurring form of the DIMIC protein by, for example, competitively modulating a cellular activity of a DIMIC protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the DIMIC protein.

In one embodiment, variants of a DIMIC protein which function as either DIMIC agonists (mimetics) or as DIMIC antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of a DIMIC protein for DIMIC protein agonist or antagonist activity. In one embodiment, a variegated library of DIMIC variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of DIMIC variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene

sequences such that a degenerate set of potential DIMIC sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of DIMIC sequences therein. There are a variety of methods which can be used to produce libraries of potential DIMIC variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential DIMIC sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of a DIMIC protein coding sequence can be used to generate a variegated population of DIMIC fragments for screening and subsequent selection of variants of a DIMIC protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a DIMIC coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the DIMIC protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of DIMIC proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the

screening assays to identify DIMIC variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated DIMIC library. For example, a library of expression vectors can be transfected into a cell line which ordinarily synthesizes and secretes DIMIC. The transfected cells are then cultured such that DIMIC and a particular mutant DIMIC are secreted and the effect of expression of the mutant on DIMIC activity in cell supernatants can be detected, *e.g.*, by any of a number of enzymatic assays. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of DIMIC activity, and the individual clones further characterized.

An isolated DIMIC protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind DIMIC using standard techniques for polyclonal and monoclonal antibody preparation. A full-length DIMIC protein can be used or, alternatively, the invention provides antigenic peptide fragments of DIMIC for use as immunogens. The antigenic peptide of DIMIC comprises at least 8 amino acid residues and encompasses an epitope of DIMIC such that an antibody raised against the peptide forms a specific immune complex with DIMIC. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of DIMIC that are located on the surface of the protein, *e.g.*, hydrophilic regions.

A DIMIC immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed DIMIC protein or a chemically synthesized DIMIC polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic DIMIC preparation induces a polyclonal anti-DIMIC antibody response.

Accordingly, another aspect of the invention pertains to anti-DIMIC antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as DIMIC. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with

an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind DIMIC. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of DIMIC. A monoclonal antibody composition thus typically displays a single binding affinity for a particular DIMIC protein with which it immunoreacts.

Polyclonal anti-DIMIC antibodies can be prepared as described above by immunizing a suitable subject with a DIMIC immunogen. The anti-DIMIC antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized DIMIC. If desired, the antibody molecules directed against DIMIC can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-DIMIC antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem* .255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a DIMIC immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds DIMIC.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-DIMIC monoclonal antibody (see, *e.g.*, G. Galfre *et al.* (1977) *Nature* 266:55052; Gefter *et al.* *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will

appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind DIMIC, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-DIMIC antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with DIMIC to thereby isolate immunoglobulin library members that bind DIMIC. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrad *et al.* (1991)

Bio/Technology 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al.* *Nature* (1990) 348:552-554.

Additionally, recombinant anti-DIMIC antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-DIMIC antibody (*e.g.*, monoclonal antibody) can be used to isolate DIMIC by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-DIMIC antibody can facilitate the purification of natural DIMIC from cells and of recombinantly produced DIMIC expressed in host cells. Moreover, an anti-DIMIC antibody can be used to detect DIMIC protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the DIMIC protein. These antibodies can also be used, for example, for the immunoprecipitation and immunolocalization of proteins according to the invention as well as for the monitoring of the synthesis of such proteins, for example, in recombinant organisms, and for the identification of compounds interacting with the protein according to the invention.

Anti-DIMIC antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials,

bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

"Homologues" or "Homologs" of a protein of the invention are those peptides, oligopeptides, polypeptides, proteins and enzymes which contain amino acid substitutions, deletions and/or additions relative to the said protein with respect to which they are a homologue without altering one or more of its functional properties, in particular without reducing the activity of the resulting product. For example, a homologue of said protein will consist of a bioactive amino acid sequence variant of said protein. To produce such homologues, amino acids present in the said protein can be replaced by other amino acids having similar properties, for example hydrophobicity, hydrophilicity, hydrophobic moment, antigenicity, propensity to form or break α-helical structures or β-sheet structures, and so on. Two special forms of homology, orthologous and paralogous, are evolutionary concepts used to describe ancestral relationships of genes. The term "paralogous" relates to gene-duplications within the genome of a species leading to paralogous genes. The term "orthologous" relates to homologous genes in different organisms due to ancestral relationship. The present invention thus also relates to homologues, paralogues and orthologues of the genes and proteins of the invention. The paralogues or orthologues of the genes and proteins of the invention may have a lesser percentage of sequence identity with the sequences or proteins of the invention than the strictly interpreted "homologues" as defined earlier.

"Derivatives" of a protein of the invention are those peptides, oligopeptides, polypeptides, proteins and enzymes which may comprise additional naturally-occurring, altered glycosylated, acylated or non-naturally occurring amino acid residues compared to the amino acid sequence of a naturally-occurring form of said polypeptide. Alternatively or in addition, a derivative may comprise one or more non-amino acid substituents compared to the amino acid sequence of a naturally-occurring form of said polypeptide, for example a reporter molecule or other ligand, covalently or non-covalently bound to the amino acid sequence such as, for example, a reporter molecule which is bound thereto to

facilitate its detection. A derivative of a protein retains the biological or enzymatical activity of the protein where it is derived from.

III. Computer Readable Means

The nucleotide or amino acid sequences of the invention are also provided in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a nucleotide or amino acid sequences of the present invention. Such a manufacture provides the nucleotide or amino acid sequences, or a subset thereof (*e.g.*, a subset of open reading frames (ORI's)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exist in nature or in purified form.

In one application of this embodiment, a nucleotide or amino acid sequence of the present invention can be recorded on computer readable media. As used herein "computer readable media" includes any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such a CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled artisan will readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention.

As used herein "recorded" refers to a process of storing information on computer readable medium. The skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising the nucleotide or amino acid sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form

of an ASCII file, stored in a database application, such as DB2, Sybase Oracle, or the like. The skilled artisan can readily adapt any number of dataprocessor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identity fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotide or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or form about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software of conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPatter (EMBL), BLASTN and BASTX (NCBIA).

For example, software which implements the BLAST (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410) and BLAZE (Brutlag *et al.* (1993) *Comp. Chem.* 17:203-207) search

algorithms on a Sybase system can be used to identify open reading frames (ORFs) of the sequences of the invention which contain homology to ORFs or proteins from other libraries. Such ORFs are protein encoding fragments and are useful in producing commercially important proteins such as enzyme used in various reactions and in the production of commercially useful metabolites.

IV. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a DIMIC protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, *e.g.*, a plant cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory

sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, DIMIC proteins, mutant forms of DIMIC proteins, fusion proteins, and the like).

The vectors of the invention comprise a selectable and/or scorable marker. Selectable marker genes useful for the selection of transformed plant cells, callus, plant tissue and plants are well known to those skilled in the art and comprise, for example, antimetabolite resistance as the basis of selection for dhfr, which confers resistance to methotrexate (Reiss, Plant Physiol. (Life Sci. Adv.) 13 (1994), 143-149); npt, which confers resistance to the aminoglycosides neomycin, kanamycin and paromycin (Herrera-Estrella, *EMBO J.* 2 (1983), 987-995) and hygro, which confers resistance to hygromycin (Marsh, *Gene* 32 (1984), 481-485). Additional selectable genes have been described, namely trpB, which allow cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine (Hartman, *Proc. Natl. Acad. Sci. USA* 85 (1988), 8047); mannose-6-phosphate isomerase which allows cells to utilize mannose (WO 94/20627) and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue, 1987, In: Current Communications in Molecular Biology, Cold Spring Harbor Laboratory ed.) or deaminase from Aspergillus terreus which confers resistance to Blasticidin S (Tamura, *Biosci. Biotechnol. Biochem.* 59 (1995), 2336-2338).

Useful scorable markers are also known to those skilled in the art and are commercially available. Advantageously, the marker is a gene encoding luciferase (Giacomin, *Pl. Sci.* 116 (1996), 59-72; Scikantha, *J. Bact.* 178 (1996), 121), green fluorescent protein (Gerdes, *FEBS Lett.* 389 (1996), 44-47) or ß-glucuronidase (Jefferson, *EMBO J.* 6 (1987), 3901-3907). This embodiment is particularly useful for

simple and rapid screening of cells, tissues and organisms containing a vector of the invention.

A "plant promoter" is a promoter capable of initiating transcription in plant cells. Exemplary plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria. Preferred promoters may contain additional copies of one or more specific regulatory elements, to further enhance expression and/or to alter the spatial expression and/or temporal expression of a nucleic acid molecule to which it is operably connected. For example, copper-responsive, glucocorticoid-responsive or dexamethasone-responsive regulatory elements may be placed adjacent to a heterologous promoter sequence driving expression of a nucleic acid molecule to confer copper inducible, glucocorticoid-inducible, or dexamethasone-inducible expression respectively, on said nucleic acid molecule. Examples of promoters under developmental control include promoters that preferentially initiate transcription in certain tissues, such as leaves, roots, seeds, endosperm, embryos, fibers, xylem vessels, tracheids, or sclerenchyma. Such promoters are referred to as "tissue preferred." Promoters which initiate transcription only in certain tissue are referred to as "tissue specific." A "cell type" specific promoter primarily drives expression in certain cell types in one or more organs, for example, vascular cells in roots or leaves. An "inducible" promoter is a promoter which is under environmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions or the presence of light. Tissue specific, tissue preferred, cell type specific, and inducible promoters constitute the class of "non-constitutive" promoters. A "constitutive" promoter is a promoter which is active under most environmental conditions.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a DIMIC protein can be expressed in plant cells, bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

Means for introducing a recombinant expression vector of this invention into plant tissue or cells include, but are not limited to, transformation using $CaCl_2$ and variations thereof, in particular the method described by Hanahan (J. Mol.Biol. 166, 557-560, 1983), direct DNA uptake into protoplasts (Krens *et al*, Nature 296: 72-74, 1982; Paszkowski *et al*, EMBO J. 3:2717-2722, 1984), PEG-mediated uptake to protoplasts (Armstrong *et al*, Plant Cell Reports 9: 335-339, 1990) microparticle bombardment, electroporation (Fromm *et al.*, Proc. Natl. Acad. Sci. (USA) 82:5824-5828, 1985), microinjection of DNA (Crossway *et al.*, Mol. Gen. Genet. 202:179-185, 1986), microparticle bombardment of tissue explants or cells (Christou *et al*, Plant Physiol 87: 671-674, 1988; Sanford, Particulate Science and Technology 5: 27-37, 1987), vacuum-infiltration of tissue with nucleic acid, or in the case of plants, T-DNA-mediated transfer from Agrobacterium to the plant tissue as described essentially by An *et al.*(EMBO J 4:277-284, 1985), Herrera-Estrella *et al.* (Nature 303: 209-213, 1983a; EMBO J. 2: 987-995, 1983b; In: Plant Genetic Engineering, Cambridge University Press, N.Y., pp 63-93, 1985), or *in planta* method using *Agrobacterium tumefaciens* such as that described by Bechtold *et al.*, (*C.R. Acad. Sci. (Paris, Sciences de la vie/ Life Sciences)316:* 1194-1199, 1993) or Clough *et al* (*Plant J. 16:* 735-743, 1998). The vector DNA may further comprise a selectable marker gene to facilitate the identification and/or selection of cells which are transfected or transformed with a genetic construct. Suitable selectable marker genes contemplated herein include the ampicillin resistance (Amp^r), tetracycline resistance gene Tc^r), bacterial kanamycin resistance gene (Kan^r), phosphinothricin resistance gene, neomycin phosphotransferase gene (*nptII*), hygromycin resistance gene, □-glucuronidase (GUS) gene, chloramphenicol acetyltransferase (CAT) gene, green fluorescent protein (*gfp*) gene (Haseloff *et al,* 1997), and luciferase gene.

Methods for transformation of monocotyledonous plants are well known in the art and include *Agrobacterium*-mediated transformation (Cheng *et al.* (1997) WO 97/48814;

Hansen (1998) WO 98/54961; Hiei *et al.* (1994) WO 94/00977; Hiei *et al.* (1998) WO 98/17813; Rikiishi *et al.* (1999) WO 99/04618; Saito *et al.* (1995) WO 95/06722), microprojectile bombardment (Adams *et al.* (1999) US 5,969,213; Bowen *et al.* (1998) US 5,736,369; Chang *et al.* (1994) WO 94/13822; Lundquist *et al.* (1999) US 5,874,265/US 5,990,390; Vasil and Vasil (1995) US 5,405,765; Walker *et al.* (1999) US 5,955,362), DNA uptake (Eval *et al.* (1993) WO 93/181,168), microinjection of *Agrobacterium* cells (von Holt 1994 DE 4309203) and sonication (Finer *et al.* (1997) US 5,693,512).

For microparticle bombardment of cells, a microparticle is propelled into a cell to produce a transformed cell. Any suitable ballistic cell transformation methodology and apparatus can be used in performing the present invention. Exemplary apparatus and procedures are disclosed by Stomp *et al.* (U.S. Patent No. 5,122,466) and Sanford and Wolf (U.S. Patent No. 4,945,050). When using ballistic transformation procedures, the gene construct may incorporate a plasmid capable of replicating in the cell to be transformed. Examples of microparticles suitable for use in such systems include 1 to 5 µm gold spheres. The DNA construct may be deposited on the microparticle by any suitable technique, such as by precipitation.

A whole plant may be regenerated from the transformed or transfected cell, in accordance with procedures well known in the art. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a gene construct of the present invention and a whole plant regenerated therefrom. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (*e.g.*, apical meristem, axillary buds, and root meristems), and induced meristem tissue (*e.g.*, cotyledon meristem and hypocotyl meristem).

The term "organogenesis", as used herein, includes a process by which shoots and roots are developed sequentially from meristematic centres.

The term "embryogenesis", as used herein, includes a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes.

Preferably, the plant is produced according to the methods of the invention by transfecting or transforming the plant with a genetic sequence, or by introducing to the plant a protein, by any art-recognized means, such as microprojectile bombardment, microinjection, *Agrobacterium*-mediated transformation (including *in planta*

transformation), protoplast fusion, or electroporation, amongst others. Most preferably the plant is produced by *Agrobacterium*-mediated transformation.

Agrobacterium-mediated transformation or agrolistic transformation of plants, yeast, moulds or filamentous fungi is based on the transfer of part of the transformation vector sequences, called the T-DNA, to the nucleus and on integration of said T-DNA in the genome of said eukaryote.

The term "*Agrobacterium*" as used herein, includes a member of the *Agrobacteriaceae*, more preferably *Agrobacterium* or *Rhizobacterium* and most preferably *Agrobacterium tumefaciens*.

The term "T-DNA", or "transferred DNA", as used herein, includes the transformation vector flanked by T-DNA borders which is, after activation of the *Agrobacterium vir* genes, nicked at the T-DNA borders and is transferred as a single stranded DNA to the nucleus of an eukaryotic cell.

As used herein, the terms "T-DNA borders", "T-DNA border region", or "border region" include either right T-DNA borders (RB) or left T-DNA borders (LB), which comprise a core sequence flanked by a border inner region as part of the T-DNA flanking the border and/or a border outer region as part of the vector backbone flanking the border. The core sequences comprise 22 bp in case of octopine-type vectors and 25 bp in case of nopaline-type vectors. The core sequences in the right border region and left border region form imperfect repeats. Border core sequences are indispensable for recognition and processing by the *Agrobacterium* nicking complex consisting of at least VirD1 and VirD2. Core sequences flanking a T-DNA are sufficient to promote transfer of the T-DNA. However, efficiency of transformation using transformation vectors carrying the T-DNA solely flanked by the core sequences is low. Border inner and outer regions are known to modulate efficiency of T-DNA transfer (Wang *et al.* 1987). One element enhancing T-DNA transfer has been characterized and resides in the right border outer region and is called *overdrive* (Peralta *et al.* 1986, van Haaren *et al.* 1987).

As used herein, the term "T-DNA transformation vector" or "T-DNA vector" includes any vector encompassing a T-DNA sequence flanked by a right and left T-DNA border consisting of at least the right and left border core sequences, respectively, and used for transformation of any eukaryotic cell.

As used herein, the term "T-DNA vector backbone sequence" or "T-DNA vector backbone sequences" includes all DNA of a T-DNA containing vector that lies outside of the T-DNA borders and, more specifically, outside the nicking sites of the border core imperfect repeats.

The present invention includes optimized T-DNA vectors such that vector backbone integration in the genome of a eukaryotic cell is minimized or absent. The term "optimized T-DNA vector" as used herein includes a T-DNA vector designed either to decrease or abolish transfer of vector backbone sequences to the genome of a eukaryotic cell. Such T-DNA vectors are known to the one of skill in the art and include those described by Hanson *et al.* (1999) and by Stuiver *et al.* (1999 - WO9901563).

The current invention clearly considers the inclusion of a DNA sequence encoding a DIMIC, homologue, analogue, derivative or immunologically active fragment thereof as defined supra, in any T-DNA vector comprising binary transformation vectors, super-binary transformation vectors, co-integrate transformation vectors, Ri-derived transformation vectors as well as in T-DNA carrying vectors used in agrolistic transformation.

As used herein, the term "binary transformation vector" includes a T-DNA transformation vector comprising: a T-DNA region comprising at least one gene of interest and/or at least one selectable marker active in the eukaryotic cell to be transformed; and

a vector backbone region comprising at least origins of replication active in *E. coli* and *Agrobacterium* and markers for selection in *E. coli* and *Agrobacterium*. Alternatively, replication of the binary transformation vector in *Agrobacterium* is dependent on the presence of a separate helper plasmid. The binary vector pGreen and the helper plasmid pSoup form an example of such a system (Hellens *et al.* (2000) *Plant Mol. Biol.* 42, 819-832; http://www.pgreen.ac.uk).

The T-DNA borders of a binary transformation vector can be derived from octopine-type or nopaline-type Ti plasmids or from both. The T-DNA of a binary vector is only transferred to a eukaryotic cell in conjunction with a helper plasmid. As used herein, the term "helper plasmid" includes a plasmid that is stably maintained in *Agrobacterium* and is at least carrying the set of *vir* genes necessary for enabling transfer of the T-DNA. The set of *vir* genes can be derived from either octopine-type or nopaline-type Ti plasmids or from both.

As used herein, the term "super-binary transformation vector" includes a binary transformation vector additionally carrying in the vector backbone region a *vir* region of the Ti plasmid pTiBo542 of the super-virulent *A. tumefaciens* strain A281 (EP0604662, EP0687730). Super-binary transformation vectors are used in conjunction with a helper plasmid.

As used herein, the term "co-integrate transformation vector" includes a T-DNA vector at least comprising: a T-DNA region comprising at least one gene of interest and/or at least one selectable marker active in plants; and a vector backbone region comprising at least origins of replication active in *Escherichia coli* and *Agrobacterium,* and markers for selection in *E. coli* and *Agrobacterium,* and a set of *vir* genes necessary for enabling transfer of the T-DNA. The T-DNA borders and the set of *vir* genes of the T-DNA vector can be derived from either octopine-type or nopaline-type Ti plasmids or from both.

The term "Ri-derived plant transformation vector" includes a binary transformation vector in which the T-DNA borders are derived from a Ti plasmid and the binary transformation vector being used in conjunction with a 'helper' Ri-plasmid carrying the necessary set of *vir* genes.

The terms "agrolistics", "agrolistic transformation" or "agrolistic transfer" include a transformation method combining features of *Agrobacterium*-mediated transformation and of biolistic DNA delivery. As such, a T-DNA containing target plasmid is co-delivered with DNA/RNA enabling in planta production of VirD1 and VirD2 with or without VirE2 (Hansen and Chilton 1996; Hansen *et al.* 1997; Hansen and Chilton 1997 - WO9712046).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a DIMIC protein. Accordingly, the invention further provides methods for producing a DIMIC protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a DIMIC protein has been introduced) in a suitable medium such that a DIMIC protein is produced. In another embodiment, the method further comprises isolating a DIMIC protein from the medium or the host cell.

The host cells of the invention can also be used to produce transgenic plant or non-human transgenic animals in which exogenous DIMIC sequences have been introduced into their genome or homologous recombinant plants or animals in which endogenous DIMIC sequences have been altered. Such plants and animals are useful for studying the function and/or activity of a DIMIC and for identifying and/or evaluating modulators of DIMIC activity.

Trangenic Plants

As used herein, "transgenic plant" includes a plant which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to

successive generations. The heteroglogous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses as asexual propagation from the initial transgenic. The term "transgenic" as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring event such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

A transgenic plant of the invention can be created by introducing a DIMIC-encoding nucleic acid into the plant by placing it under the control of regulatory elements which ensure the expression in plant cells. These regulatory elements may be heterologous or homologous with respect to the nucleic acid molecule to be expressed as well with respect to the plant species to be transformed. In general, such regulatory elements comprise a promoter active in plant cells. These promoters can be used to modulate (*e.g.* increase or decrease) DIMIC content and/or composition in a desired tissue. To obtain expression in all tissues of a transgenic plant, preferably constitutive promoters are used, such as the 35 S promoter of CaMV (Odell, Nature 313 (1985), 810-812) or promoters from such genes as rice actin (McElroy *et al.* (1990) Plant Cell 2:163-171) maize H3 histone (Lepetit *et al.* (1992) Mol. Gen. Genet 231:276-285) or promoters of the polyubiquitin genes of maize (Christensen, Plant Mol. Biol. 18 (1982), 675-689). In order to achieve expression in specific tissues of a transgenic plant it is possible to use tissue specific promoters (see, *e.g.*, Stockhaus, EMBO J. 8 (1989), 2245-2251 or Table II, below).

Table II:

I: CELL-SPECIFIC,TISSUE-SPECIFIC, AND ORGAN-SPECIFIC PROMOTERS		
GENE SOURCE	**EXPRESSION PATTERN**	**REFERENCE**
α-amylase (*Amy32b*)	aleurone	Lanahan et al, Plant Cell 4:203-211, 1992; Skriver et al, Proc Natl Acad Sci USA 88:7266-7270, 1991
cathepsin β-like gene	aleurone	Cejudo et al, Plant Mol Biol 20:849-856, 1992
Agrobacterium rhizogenes rolB	cambium	Nilsson et al, Physiol Plant 100:456-462, 1997
AtPRP4	flowers	http://salus.medium.edu/mmg/tierney/html
chalcone synthase (chsA)	flowers	Van der Meer et al, Plant Mol Biol 15:95-109, 1990
LAT52	anther	Twell et al, Mol Gen Genet 217:240-245, 1989
apetala-3	flowers	
Chitinase	fruit (berries, grapes, etc)	Thomas et al. CSIRO Plant Industry, Urrbrae, South Australia, Australia; http://winetitles.com.au/gwrdc/csh95-1.html
rbcs-3A	green tissue (eg leaf)	Lam et al, Plant Cell 2:857-866, 1990; Tucker et al., Plant Physiol 113:1303-1308, 1992
leaf-specific genes	leaf	Baszczynski et al, Nucl Acid Res 16:4732, 1988
AtPRP4	leaf	http://salus.medium.edu/mmg/tierney/html
chlorella virus adenine methyltransferase gene promoter	leaf	Mitra and Higgins, Plant Mol Biol 26:85-93, 1994
aldP gene promoter from rice	leaf	Kagaya et al, Mol Gen Genet 248:668-674, 1995
rbcs promoter from rice or tomato	leaf	Kyozuka et al, Plant Physiol 102:991-1000, 1993
Pinus cab-6	leaf	Yamamoto et al, Plant Cell Physiol 35:773-778, 1994
rubisco promoter	leaf	
cab (chlorophyll a/b/binding protein	leaf	
pea Blec4 gene	vegetative and floral epidermal tissues	Mandaci and Dobres, Plant Mol Biol 34:961-965
SAM22	senescent leaf	Crowell et al, Plant Mol Biol 18:459-466, 1992
ltp gene (lipid transfer gene)		Fleming et al, Plant J 2:855-862, 1992
R. japonicum nif gene	nodule	United States Patent No 4 803165

I: CELL-SPECIFIC,TISSUE-SPECIFIC, AND ORGAN-SPECIFIC PROMOTERS		
GENE SOURCE	**EXPRESSION PATTERN**	**REFERENCE**
B. japonicum nifH gene	nodule	United States Patent No 5008194
GmENOD40	nodule	Yang et al, Plant J 3:573-585, 1993
PEP carboxylase (PEPC)	nodule	Pathirana et al, Plant Mol Biol 20:437-450, 1992
leghaemoglobin (Lb)	nodule	Gordon et al, J Exp Bot 44:1453-1465, 1993
Tungro bacilliform virus gene	phloem	Bhattacharyya-Pakrasi et al, Plant J 4:71-79, 1992
pollen-specific genes	pollen; microspore	Albani et al, Plant Mol Biol 15:605, 1990; Albani et al, Plant Mol Biol 16:501, 1991
Zm13	pollen	Guerrero et al, Mol Gen Genet 224:161-168, 1993
apg gene	microspore	Twell et al, Sex Plant Reprod 6:217-224, 1993
maize pollen-specific gene	pollen	Hamilton et al, Plant Mol Biol 18:211-218, 1992
sunflower pollen-expressed gene	pollen	Baltz et al, Plant J 2:713-721, 1992
B. napus pollen-specific gene	pollen;anther; tapetum	Arnoldo et al, J Cell Biochem, Abstract No. Y101, 204, 1992
root-expressible genes	roots	Tingey et al, EMBO J 6:1, 1987
tobacco auxin-inducible gene	root tip	Van der Zaal et al, Plant Mol Biol 16:983, 1991
β-tubulin	root	Oppenheimer et al, Gene 63:87, 1988
tobacco root-specific genes	root	Conkling et al, Plant Physiol 93:1203, 1990
B. napus G1-3b gene	root	United States Patent No 5401836
SbPRP1	roots	Suzuki et al, Plant Mol Biol 21:109-119, 1993
AtPRP1; AtPRP3	roots; root hairs	http://salus.medium.edu/mmg/tierney/html
RD2 gene	root cortex	http://www2.cnsu.edu/ncsu/research
TobRB7 gene	root vasculature	http://www2.cnsu.edu/ncsu/research
AtPRP4	leaves; flowers; lateral root primordia	http://salus.medium.edu/mmg/tierney/html
seed-specific genes	seed	Simon et al, Plant Mol Biol 5:191, 1985; Scofield et al, J Biol Chem 262:12202, 1987; Baszczynski et al, Plant Mol Biol 14:633, 1990
Brazil Nut albumin	seed	Pearson et al, Plant Mol Biol 18:235-245, 1992
Legumin	seed	Ellis et al, Plant Mol Biol 10:203-214, 1988

I: CELL-SPECIFIC,TISSUE-SPECIFIC, AND ORGAN-SPECIFIC PROMOTERS		
GENE SOURCE	EXPRESSION PATTERN	REFERENCE
glutelin (rice)	seed	Takaiwa et al, Mol Gen Genet 208:15-22, 1986; Takaiwa et al, FEBS Lett 221:43-47, 1987
Zein	seed	Matzke et al, Plant Mol Biol 14:323-32 1990
NapA	seed	Stalberg et al, Planta 199:515-519, 1996
wheat LMW and HMW glutenin-1	endosperm	Mol Gen Genet 216:81-90, 1989; Nucl Acids Res 17:461-462, 1989
wheat SPA	seed	Albani et al, Plant Cell 9:171-184, 1997
cZ19B1, maize 19 kDa zein	seed	WO0011177
mi1ps, maize myoinositol-1-Pi synthase	seed	WO0011177
wheat α, β, γ-gliadins	endosperm	EMBO J 3:1409-1415, 1984
barley *ltr1* promoter	endosperm	
barley B1, C, D, hordein	endosperm	Theor Appl Gen 98:1253-1262, 1999; Plant J 4:343-355, 1993; Mol Gen Genet 250:750-60, 1996
barley DOF	endosperm	Mena et al, Plant J 116:53-62, 1998
blz2	endosperm	EP99106056.7
synthetic promoter	endosperm	Vicente-Carbajosa et al, Plant J 13:629-640, 1998
rice prolamin NRP33	endosperm	Wu et al, Plant Cell Physiol 39: 885-889, 1998
rice α-globulin Glb-1	endosperm	Wu et al, Plant Cell Physiol 39:885-889, 1998
maize END genes	endosperm	WO0012733
barley END1	endosperm	WO9808961
barley NUC1	nucellus	WO9808961
rice OSH1	embryo	Sato et al, Proc Natl Acad Sci USA 93:8117-8122, 1996
rice α-globulin REB/OHP-1	endosperm	Nakase et al, Plant Mol Biol 33:513-522, 1997
rice ADP-glucose PP	endosperm	Trans Res 6:157-168, 1997
maize ESR gene family	endosperm	Plant J 12:235-246, 1997
sorgum γ-kafirin	endosperm	Plant Mol Biol 32:1029-1035, 1996
KNOX	embryo	Postma-Haarsma et al, Plant Mol Biol 39:257-271, 1999
rice oleosin	embryo and aleuron	Wu et al, J Biochem 123:386, 1998
sunflower oleosin	seed (embryo and dry seed)	Cummins et al, Plant Mol Biol 19:873-876, 1992

I: CELL-SPECIFIC,TISSUE-SPECIFIC, AND ORGAN-SPECIFIC PROMOTERS		
GENE SOURCE	**EXPRESSION PATTERN**	**REFERENCE**
LEAFY	shoot meristem	Weigel et al, Cell 69:843-859, 1992
Arabidopsis thaliana knat1	shoot meristem	Accession number AJ131822
Malus domestica kn1	shoot meristem	Accession number Z71981
CLAVATA1	shoot meristem	Accession number AF049870
stigma-specific genes	stigma	Nasrallah et al, Proc Natl Acad Sci USA 85:5551, 1988; Trick et al, Plant Mol Biol 15:203, 1990
class I patatin gene	tuber	Liu et al, Plant Mol Biol 153:386-395, 1991
PCNA rice	meristem	Kosugi et al, Nucl Acids Res 19:1571-1576, 1991; Kosugi and Ohashi, Plant Cell 9:1607-1619, 1997
Pea TubA1 tubulin	Dividing cells	Stotz and Long, Plant Mol Biol 41:601-614, 1999
Arabidopsis cdc2a	cycling cells	Chung and Parish, FEBS Lett 362:215-219, 1995
Arabidopsis Rop1A	Anthers; mature pollen + pollen tubes	Li et al, Plant Physiol 118:407-417, 1998
Arabidopsis AtDMC1	Meiosis-associated	Klimyuk and Jones, Plant J 11:1-14, 1997
Pea PS-IAA4/5 and PS-IAA6	Auxin-inducible	Wong et al, Plant J 9:587-599, 1996
Pea farnesyltransferase	Meristematic tissues; phloem near growing tissues; light- and sugar-repressed	Zhou et al, Plant J 12:921-930,1997
Tobacco (*N. sylvestris*) cyclin B1;1	Dividing cells / meristematic tissue	Trehin et al, Plant Mol.Biol. 35:667-672, 1997
Catharanthus roseus Mitotic cyclins CYS (A-type) and CYM (B-type)	Dividing cells / meristematic tissue	Ito et al, Plant J 11:983-992, 1997
Arabidopsis cyc1At (=cyc B1;1) and cyc3aAt (A-type)	Dividing cells / meristematic tissue	Shaul et al, Proc Natl Acad Sci USA 93:4868-4872, 1996
Arabidopsis tef1 promoter box	Dividing cells / meristematic tissue	Regad et al, Mol Gen Genet 248:703-711, 1995
Catharanthus roseus cyc07	Dividing cells / meristematic tissue	Ito et al, Plant Mol Biol 24:863-878, 1994

II: CONSTITUTIVE PROMOTERS		
GENE SOURCE	**EXPRESSION PATTERN**	**REFERENCE**
Actin	constitutive	McElroy et al, Plant Cell 2:163-171, 1990
CAMV 35S	constitutive	Odell et al, Nature 313:810-812, 1985
CaMV 19S	constitutive	Nilsson et al, Physiol Plant 100:456-462, 1997
GOS2	constitutive	de Pater et al, Plant J 2:837-844, 1992
Ubiquitin	constitutive	Christensen et al, Plant Mol Biol 18:675-689, 1992
rice cyclophilin	constitutive	Buchholz et al, Plant Mol Biol 25:837-843, 1994
maize histone H3	constitutive	Lepetit et al, Mol Gen Genet 231:276-285, 1992
alfalfa histone H3	constitutive	Wu et al, Nucleic Acids Res 17:3057-3063, 1989; Wu et al, Plant Mol Biol 11:641-649, 1988
actin 2	constitutive	An et al, Plant J 10:107-121, 1996
III: STRESS-INDUCIBLE PROMOTERS		
GENE SOURCE	**EXPRESSION PATTERN**	**REFERENCE**
P5CS (delta(1)-pyrroline-5-carboxylate syntase)	salt, water	Zhang et al, Plant Sci 129:81-89, 1997
cor15a	cold	Hajela et al, Plant Physiol 93:1246-1252, 1990
cor15b	cold	Wlihelm et al, Plant Mol Biol 23:1073-1077, 1993
cor15a (-305 to +78 nt)	cold, drought	Baker et al, Plant Mol Biol 24: 01-713, 1994
rd29	salt, drought, cold	Kasuga et al, Nature Biotechnol 18:287-291, 1999
heat shock proteins, including artificial promoters containing the heat shock element (HSE)	heat	Barros et al, Plant Mol Biol 19 665-75, 1992. Marrs et al, Dev Genet14:27-41, 1993. Schoffl et al, Mol Gen Genet 217:246-53, 1989.
smHSP (small heat shock proteins)	heat	Waters et al, J Exp Bot 47:325-338, 1996
wcs120	cold	Ouellete et al, FEBS Lett 423:324-328, 1998
ci7	cold	Kirch et al, Plant Mol Biol 33:897-909, 1997
Adh	cold, drought, hypoxia	Dolferus et al, Plant Physiol 105:1075-87, 1994
pwsi18	water: salt and drought	Joshee et al, Plant Cell Physiol 39:64-72, 1998

III: STRESS-INDUCIBLE PROMOTERS		
GENE SOURCE	**EXPRESSION PATTERN**	**REFERENCE**
ci21A	cold	Schneider et al, Plant Physiol 113:335-45, 1997
Trg-31	drought	Chaudhary et al, Plant Mol Biol 30:1247-57, 1996
Osmotin	osmotic	Raghothama et al, Plant Mol Biol 23:1117-28, 1993
LapA	wounding, enviromental	WO99/03977 University of California/INRA
IV: PATHOGEN-INDUCIBLE PROMOTERS		
GENE SOURCE	**EXPRESSION PATTERN**	**REFERENCE**
RB7	Root-knot nematodes (Meloidogyne spp.)	US5760386 - North Carolina State University; Opperman et al, Science 263:221-23, 1994
PR-1, 2, 3, 4, 5, 8, 11	fungal, viral, bacterial	Ward et al, Plant Cell 3:1085-1094, 1991; Reiss et al 1996; Lebel et al, Plant J 16:223-233, 1998; Melchers et al, Plant J 5:469-480, 1994; Lawton et al, Plant Mol Biol, 19:735-743, 1992
HMG2	nematodes	WO9503690 - Virginia Tech Intellectual Properties Inc .
Abi3	Cyst nematodes (Heterodera spp.)	unpublished
ARM1	nematodes	Barthels et al, Plant Cell 9:2119-2134, 1997 WO 98/31822 – Plant Genetic Systems
Att0728	nematodes	Barthels et al, Plant Cell 9: 2119-2134, 1997 PCT/EP98/07761
Att1712	nematodes	Barthels et al, Plant Cell 9, 2119-2134, 1997 PCT/EP98/07761
Gst1	Different types of pathogens	Strittmatter et al, Mol Plant-Microbe Interact 9:68-73, 1996
LEMMI	nematodes	WO 92/21757 – Plant Genetic Systems
CLE	geminivirus	PCT/EP99/03445 - CINESTAV
PDF1.2	Fungal including *Alternaria brassicicola* and *Botrytis cinerea*	Manners et al, Plant Mol Biol, 38:1071-1080, 1998
Thi2.1	Fungal – *Fusarium oxysporum f sp. matthiolae*	Vignutelli et al, Plant J 14:285-295, 1998
DB#226	nematodes	Bird and Wilson, Mol Plant-Microbe Interact 7:419-442, 1994 WO 95.322888

IV: PATHOGEN-INDUCIBLE PROMOTERS		
GENE SOURCE	EXPRESSION PATTERN	REFERENCE
DB#280	nematodes	Bird and Wilson, Mol Plant-Microbe Interact 7:419-442, 1994 WO 95.322888
Cat2	nematodes	Niebel et al, Mol Plant-Microbe Interact 8:371-378, 1995
□Tub	nematodes	Aristizabal et al (1996), 8th International Congress on Plant-Microbe Interaction, Knoxville US B-29
SHSP	nematodes	Fenoll et al (1997) In: Cellular and molecular aspects of plant-nematode interactions. Kluwer Academic, C. Fenoll, F.M.W. Grundler and S.A. Ohl (Eds.),
Tsw12	nematodes	Fenoll et al (1997) In: Cellular and molecular aspects of plant-nematode interactions. Kluwer Academic, C. Fenoll, F.M.W. Grundler and S.A. Ohl (Eds.)
Hs1(pro1)	nematodes	WO 98/122335 - Jung
NsLTP	viral, fungal, bacterial	Molina and Garcia-Olmedo FEBS Lett, 316:119-122, 1993
RIP	viral, fungal	Tumer et al, Proc Natl Acad Sci USA 94:3866-3871, 1997

The promoters listed in the foregoing table are provided for the purposes of exemplification only and the present invention is not to be limited by the list provided therein. Those skilled in the art will readily be in a position to provide additional promoters that are useful in performing the present invention. The promoters listed may also be modified to provide specificity of expression as required.

Known are also promoters which are specifically active in tubers of potatoes or in seeds of different plants species, such as maize, Vicia, wheat, barley and the like. Inducible promoters may be used in order to be able to exactly control expression under certain environmental or developmental conditions such as pathogens, anaerobia, or light. Examples of inducible promoters include the promoters of genes encoding heat shock proteins or microspore-specific regulatory elements (WO96/16182). Furthermore, the chemically inducible Tet-system may be employed (Gatz, *Mol. Gen. Genet.* 227 (1991); 229-237). Further suitable promoters are known to the person skilled in the art and are described, *e.g.*, in Ward (Plant Mol. Biol. 22 (1993), 361-366). The regulatory elements may further comprise transcriptional and/or translational enhancers functional in plants cells. Furthermore, the regulatory elements may include transcription termination signals, such as

a poly-A signal, which lead to the addition of a poly A tail to the transcript which may improve its stability.

In the case that a nucleic acid molecule according to the invention is expressed in the sense orientation, the coding sequence can be modified such that the protein is located in any desired compartment of the plant cell, *e.g.*, the nucleus, endoplasmatic reticulum, the vacuole, the mitochondria, the plastids, the apoplast, or the cytoplasm.

Methods for the introduction of foreign DNA into plants are also well known in the art. These include, for example, the transformation of plant cells or tissues with T-DNA using Agrobacterium tumefaciens or Agrobacterium rhizogenes, the fusion of protoplasts, direct gene transfer (see, *e.g.*, EP-A 164 575), injection, electroporation, biolistic methods like particle bombardment, pollen-mediated transformation, plant RNA virus-mediated transformation, liposome-mediated transformation, transformation using wounded or enzyme-degraded immature embryos, or wounded or enzyme-degraded embryogenic callus and other methods known in the art. The vectors used in the method of the invention may contain further functional elements, for example "left border"- and "right border"-sequences of the T-DNA of Agrobacterium which allow for stably integration into the plant genome. Furthermore, methods and vectors are known to the person skilled in the art which permit the generation of marker free transgenic plants, *i.e.*, the selectable or scorable marker gene is lost at a certain stage of plant development or plant breeding. This can be achieved by, for example, cotransformation (Lyznik, Plant Mol. Biol. 13 (1989), 151-161; Peng, Plant Mol. Biol. 27 (1995), 91-104) and/or by using systems which utilize enzymes capable of promoting homologous recombination in plants (see, *e.g.*, WO97/08331; Bayley, Plant Mol. Biol. 18 (1992), 353-361); Lloyd, Mol. Gen. Genet. 242 (1994), 653-657; Maeser, Mol. Gen. Genet. 230 (1991), 170-176; Onouchi, Nucl. Acids Res. 19 (1991), 6373-6378). Methods for the preparation of appropriate vectors are described by, *e.g.*, Sambrook (Molecular Cloning; A Laboratory Manual, 2nd Edition (1989), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Suitable strains of Agrobacterium tumefaciens and vectors, as well as transformation of Agrobacteria, and appropriate growth and selection media are described in, for example, GV3101 (pMK90RK), Koncz, Mol. Gen. Genet. 204 (1986), 383-396; C58C1 (pGV 3850kan), Deblaere, Nucl. Acid Res. 13 (1985), 4777; Bevan, Nucleic. Acid Res. 12(1984), 8711; Koncz, Proc. Natl. Acad. Sci. USA 86 (1989), 8467-8471; Koncz, Plant Mol. Biol. 20 (1992), 963-976; Koncz, Specialized vectors for gene tagging and expression studies. In: Plant Molecular Biology Manual Vol 2, Gelvin and Schilperoort (Eds.), Dordrecht, The Netherlands: Kluwer Academic Publ. (1994), 1-22; EP-

A-120 516; Hoekema: The Binary Plant Vector System, Offsetdrukkerij Kanters B.V., Alblasserdam (1985), Chapter V, Fraley, Crit. Rev. Plant. Sci., 4, 1-46; An, EMBO J. 4 (1985), 277-287). Although the use of Agrobacterium tumefaciens is preferred in the method of the invention, other Agrobacterium strains, such as Agrobacterium rhizogenes, may be used, for example, if a phenotype conferred by said strain is desired.

Methods for the transformation using biolistic methods are known to the person skilled in the art; see, *e.g.*, Wan, Plant Physiol. 104 (1994), 37-48; Vasil, Bio/Technology 11 (1993), 1553-1558 and Christou (1996) Trends in Plant Science 1, 423-431. Microinjection can be performed as described in Potrykus and Spangenberg (eds.), Gene Transfer To Plants. Springer Verlag, Berlin, NY (1995).

The transformation of most dicotyledonous plants may be performed using the methods described above or using transformation via biolistic methods as, *e.g.*, described above as well as protoplast transformation, electroporation of partially permeabilized cells, or introduction of DNA using glass fibers.

In general, the plants which are modified according to the invention may be derived from any desired plant species. They can be monocotyledonous plants or dicotyledonous plants, preferably they belong to plant species of interest in agriculture, wood culture or horticulture interest, such as crop plants (*e.g.*, maize, rice, barley, wheat, rye, oats), potatoes, oil producing plants (*e.g.*, oilseed rape, sunflower, pea nut, soy bean), cotton, sugar beet, sugar cane, leguminous plants (*e.g.*, beans, peas), or wood producing plants, preferably trees.

The present invention also relates to a transgenic plant cell which contains (preferably stably integrated into its genome) a nucleic acid molecule of the present invention linked to regulatory elements which allow expression of the nucleic acid molecule in plant cells. The presence and expression of the nucleic acid molecule in the transgenic plant cells leads to the synthesis of a DIMIC protein and may lead to physiological and phenotypic changes in plants containing such cells.

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype. Such regeneration techniques often rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced with a polynucleotide of the present invention.

Plant cells transformed with a plant expression vector can be regenerated, *e.g.*, from single cells, callus tissue or leaf discs according to standard plant tissue culture

techniques. It is well known in the art that various cells, tissues, and organs from almost any plant can be successfully cultured to regenerate an entire plant. Plant regeneration from cultured protoplasts is described in Evans *et al.*, *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, Macmillilan Publishing Company, New York, pp. 124-176 (1983); and *Binding, Regeneration of Plants, Plant Protoplasts*, CRC Press, Boca Raton, pp. 21-73 (1985).

Transformed plant cells, calli or explant can be cultured on regeneration medium in the dark for several weeks, generally about 1 to 3 weeks to allow the somatic embryos to mature. Preferred regeneration media include media containing MS salts, such as PHI-E and PHI-F media. The plant cells, calli or explant are then typically cultured on rooting medium in a light/dark cycle until shoots and roots develop. Methods for plant regeneration are known in the art and preferred methods are provided by Kamo *et al.*, (*Bot. Gaz.* 146(3):324-334, 1985), West *et al.*, (*The Plant Cell* 5:1361-1369. 1993), and Duncan *et al.* (*Planta* 165:322-332, 1985).

Small plantlets can then be transferred to tubes containing rooting medium and allowed to grow and develop more roots for approximately another week. The plants can then be transplanted to soil mixture in pots in the greenhouse.

The regeneration of plants containing the foreign gene introduced by *Agrobacterium* from leaft explants can be achieved as described by Horsch *et al.*, *Science*, 227:1229-1231 (1985). In this procedure, transformants are grown in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant species being transformed as described by Fraley *et al., Proc. Natl. Acad. Sci, U.S.A.* 80:4803 (1983). This procedure typically produces shoots within two to four weeks and these transformant shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Transgenic plants of the present invention may be fertile or sterile.

Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al., Ann. Rev. of Plant Phys.*, 38:467-486(1987). The regeneration of plants from either single plant protoplasts or various explants is well known in the art. See, from example, *Methods for Plant Molecular Biology*, A. Weissbach and H. Weissback, eds., Academic Press, Inc., San Diego, Calif. (1988). This regeneration and growth process includes the steps of selection of transformant cells and shoots, rooting ht transformant shoots and growth of the plantlets in soil. For maize cell culture and regeneration see generally, *The Maize Handbook*, Freeling and Walbot, Eds., Springer, New York (1994); *Corn and Corn*

Improvement, 3rd edition, Sprague and Dudley Eds., American Society of Agronomy, Madison, Wisconsin (1988).

One of skill will recognize that after the recombinant expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

In vegetatively propagated crops, mature transgenic plants can be propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. Selection of desirable transgenics is made and new varieties are obtained and propagated vegetatively for commercial use. In seed propagated crops, mature transgenic plants can be self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the newly introduced heterologous nucleic acid. These seeds can be grown to produce plants that would produce the selected phenotype, (*e.g.*, altered cell cycle content or composition).

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit and the like are included in the invention, provided that these parts comprise cells comprising the isolated nucleic acid of the present invention. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced nucleic acid sequences.

Transgenic plants expressing the selectable marker can be screened for transmission of the nucleic acid of the present invention by, for example, standard immunoblot and DNA detection techniques. Transgenic lines are also typically evaluated on levels of expression of the heterologous nucleic acid. Expression at the RNA level can be determined initially to identify and quantitate expression-positive plants. Standard techniques for RNA analysis can be employed and include PCR amplification assays using oligonucleotide primers designed to amplify only the heterologous RNA templates and solution hybridization assays using heterologous nucleic acid-specific probes. The RNA-positive plants can then analyzed for protein expression by Western immunoblot analysis using the specifically reactive antibodies of the present invention. In addition, in situ hybridization and immunocytochemistry according to standard protocols can be done using heterologous nucleic acid specific polynucleotide probes and antibodies, respectively, to localize sites of expression within transgenic tissue. Generally, a number of transgenic lines are usually screened for the incorporated nucleic acid to identify and select plants with the most appropriate expression profiles.

A preferred embodiment of the invention is a transgenic plant that is homozygous for the added heterologous nucleic acid; *i.e.*, a transgenic plant that contains two added nucleic acid sequences, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) a heterozygous transgenic plant that contains a single added heterologous nucleic acid, germinating some of the seed produced and analyzing the resulting plants produced for altered cell division relative to a control plant (*i.e.*, native, non-transgenic). Back-crossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated.

The present invention also relates to transgenic plants and plant tissue comprising transgenic plant cells according to the invention. Due to the (over)expression of a DIMIC molecule, *e.g.*, at developmental stages and/or in plant tissue in which they do not naturally occur, these transgenic plants may show various physiological, developmental and/or morphological modifications in comparison to wild-type plants.

Therefore, part of this invention is the use of the DIMIC molecules to modulate the cell cycle and/or plant cell division and/or growth in plant cells, plant tissues, plant organs and/or whole plants. To the scope of the invention also belongs a method for influencing the activity of CDKs such as CDC2a, or CDC2b, CKSs, CKIs, PLPs and KLPNTs in a plant cell by transforming the plant cell with a nucleic acid molecule according to the invention and/or manipulation of the expression of the molecule.

Furthermore, the invention also relates to a transgenic plant cell which contains (preferably stably integrated into its genome) a nucleic acid molecule of the invention or part thereof, wherein the transcription and/or expression of the nucleic acid molecule or part thereof leads to reduction of the synthesis of a DIMIC. In a preferred embodiment, the reduction is achieved by an anti-sense, sense, ribozyme, co-suppression and/or dominant mutant effect. The reduction of the synthesis of a protein according to the invention in the transgenic plant cells can result in an alteration in, *e.g.*, cell division. In transgenic plants comprising such cells this can lead to various physiological, developmental and/or morphological changes.

In yet another aspect, the invention relates to harvestable parts and to propagation material of the transgenic plants of the invention which either contain transgenic plant cells expressing a nucleic acid molecule according to the invention or which contain cells which show a reduced level of the described protein. Harvestable parts can be in principle any useful parts of a plant, for example, flowers, pollen,

seedlings, tubers, leaves, stems, fruit, seeds, roots etc. Propagation material includes, for example, seeds, fruits, cuttings, seedlings, tubers, rootstocks, and the like.

Transgenic Animals

As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous DIMIC gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a DIMIC-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The DIMIC cDNA sequence of SEQ ID NOs 35-48 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human DIMIC gene, such as a mouse or rat DIMIC gene, can be used as a transgene. Alternatively, a DIMIC gene homologue, such as another DIMIC family member, can be isolated based on hybridization to the DIMIC cDNA sequences of SEQ ID NOs 35-48 (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a DIMIC transgene to direct expression of a DIMIC protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A

transgenic founder animal can be identified based upon the presence of a DIMIC transgene in its genome and/or expression of DIMIC mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a DIMIC protein can further be bred to other transgenic animals carrying other transgenes.

V. Agricultural, Phytopharmaceutical, and Pharmaceutical Compositions

The DIMIC nucleic acid molecules, DIMIC proteins, and anti-DIMIC antibodies (also referred to herein as "active compounds") of the invention can be incorporated into compositions useful in agriculture and in plant cell and tissue culture. Plant protection compositions can be prepared by conventional means commonly used for the application of, for example, herbicides and pesticides. For example, certain additives known to those skilled in the art stabilizers or substances which facilitate the uptake by the plant cell, plant tissue or plant may be used.

The DIMIC nucleic acid molecules, DIMIC proteins, and anti-DIMIC antibodies (also referred to herein as "active compounds") of the invention can also be incorporated into pharmaceutical compositions suitable for administration into animals. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a plant or subject by, for example, injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The agricultural or pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the

agricultural or pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The agricultural and pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

VI. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) agricultural uses (*e.g.*, to increase plant yield and to develop phytopharmaceuticals); b) screening assays; c) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials); d) methods of treatment (*e.g.*, phytotherapeutic, therapeutic and prophylactic); e) transcriptomics; f) proteomics; g) metabolomics; h) ligandomics; and i) pharmacogenetics or pharmacogenomics. The isolated nucleic acid molecules of the invention can be used, for example, to express a DIMIC protein (*e.g.*, via a recombinant expression vector in a host cell or in gene therapy applications), to detect DIMIC mRNA (*e.g.*, in a biological sample) or a genetic alteration in a DIMIC gene, and to modulate DIM or DIMIC activity, as described further below. The DIMIC proteins can be used to treat disorders characterized by insufficient or excessive production of a DIMIC substrate or production of DIMIC inhibitors. In addition, the DIMIC proteins can be used to screen for naturally occurring DIMIC substrates, to screen for drugs or compounds which modulate DIMIC activity, as well as to treat disorders characterized by insufficient or excessive production of DIMIC protein or production of DIMIC protein forms which have decreased or aberrant activity compared to DIMIC wild type protein. Moreover, the anti-DIMIC antibodies of the invention can be used to detect and isolate DIMIC proteins, regulate the bioavailability of DIMIC proteins, and modulate DIMIC activity.

A. Agricultural Uses:

In another embodiment of the invention, a method is provided for modifying cell fate and/or plant development and/or plant morphology and/or plant biochemistry and/or plant physiology by modifying the expression of a DIMIC molecule of the present invention in particular cells, tissues or organs of a plant.

Modulation of the expression of a DIMIC molecule of the present invention in a plant can produce a range of desirable phenotypes in plants, such as, for example, the modification of one or more morphological, biochemical, or physiological characteristics including: (i) modification of the length of the G1 and/or the S and/or the G2 and/or the M

phase of the cell cycle of a plant; (ii) modification of the G1/S and/or S/G2 and/or G2/M and/or M/G1 phase transition of a plant cell; (iii) modification of the initiation, promotion, stimulation or enhancement of cell division; (iv) modification of the initiation, promotion, stimulation or enhancement of DNA replication;(v) modification of the initiation, promotion, stimulation or enhancement of seed set and/or seed size and/or seed development; (vi) modification of the initiation, promotion, stimulation or enhancement of tuber formation; (vii) modification of the initiation, promotion, stimulation or enhancement of fruit formation; (viii) modification of the initiation, promotion, stimulation or enhancement of leaf formation; (ix) modification of the initiation, promotion, stimulation or enhancement of shoot initiation and/or development; (x) modification of the initiation, promotion, stimulation or enhancement of root initiation and/or development; (xi) modification of the initiation, promotion, stimulation or enhancement of lateral root initiation and/or development; (xii) modification of the initiation, promotion, stimulation or enhancement of nodule formation and/or nodule function; (xiii) modification of the initiation, promotion, stimulation or enhancement of the bushiness of the plant; (xiv) modification of the initiation, promotion, stimulation or enhancement of dwarfism in the plant; (xv) modification of the initiation, promotion, stimulation or enhancement of senescence; (xvi) modification of stem thickness and/or strength characteristics and/or wind-resistance of the stem and/or stem length; (xvii) modification of tolerance and/or resistance to biotic stresses such as pathogen infection; (xviii) modification of tolerance and/or resistance to abiotic stresses such as drought stress or salt stress; (xviv) modification of the initiation, promotion, stimulation or enhancement of pre-mRNA processing; and (xx) modification of the initiation, promotion, stimulation or enhancement of vesicle trafficking/processing.

Methods to effect expression of a DIM and/or DIMIC or a homologue, analogue or derivative thereof as defined in the present invention in a plant cell, tissue or organ, include either the introduction of the protein directly to a cell, tissue or organ such as by microinjection of ballistic means or, alternatively, introduction of an isolated nucleic acid molecule encoding the protein into the cell, tissue or organ in an expressible format. Methods to effect expression of a DIM and/or DIMIC or a homologue, analogue or derivative thereof as defined in the current invention in whole plants include regeneration of whole plants from the transformed cells in which an isolated nucleic acid molecule encoding the protein was introduced in an expressible format.

The present invention clearly extends to any plant produced by the inventive method described herein, and any and all plant parts and propagules thereof. The present invention extends further to encompass the progeny derived from a primary

transformed or transfected cell, tissue, organ or whole plant that has been produced by the inventive method, the only requirement being that the progeny exhibits the same genotypic and/or phenotypic characteristic(s) as those characteristic(s) that (have) been produced in the parent by the performance of the inventive method.

Exploiting plant DIM and/or DIMIC functions to regulate plant growth and development can depend on methods comprising enhancing a DIM and/or DIMIC gene expression or ectopic expression of a DIM and/or DIMIC genes.

As used herein, the terms "ectopic expression" or "ectopic overexpression" of a gene or a protein refer to expression patterns and/or expression levels of the gene or protein normally not occurring under natural conditions.

By "cell fate and/or plant development and/or plant morphology and/or biochemistry and/or physiology" is meant that one or more developmental and/or morphological and/or biochemical and/or physiological characteristics of a plant is altered by the performance of one or more steps pertaining to the invention described herein.

"Cell fate" includes the cell-type or cellular characteristics of a particular cell that are produced during plant development or a cellular process therefor, in particular during the cell cycle or as a consequence of a cell cycle process.

The term "plant development" or the term "plant developmental characteristic" or similar terms shall, when used herein, be taken to mean any cellular process of a plant that is involved in determining the developmental fate of a plant cell, in particular the specific tissue or organ type into which a progenitor cell will develop. Cellular processes relevant to plant development will be known to those skilled in the art. Such processes include, for example, morphogenesis, photomorphogenesis, shoot development, root development, vegetative development, reproductive development, stem elongation, flowering, and regulatory mechanisms involved in determining cell fate, in particular a process or regulatory process involving the cell cycle.

The term "plant morphology" or the term "plant morphological characteristic" or similar term will, when used herein, be understood by those skilled in the art to include the external appearance of a plant, including any one or more structural features or combination of structural features thereof. Such structural features include the shape, size, number, position, color, texture, arrangement, and patternation of any cell, tissue or organ or groups of cells, tissues or organs of a plant, including the root, stem, leaf, shoot, petiole, trichome, flower, petal, stigma, style, stamen, pollen, ovule, seed, embryo, endosperm, seed coat, aleurone, fiber, fruit, cambium, wood, heartwood, parenchyma, aerenchyma, sieve element, phloem or vascular tissue.

The term "plant biochemistry" or the term "plant biochemical characteristic" or similar term will, when used herein, be understood by those skilled in the art to include the metabolic and catalytic processes of a plant, including primary and secondary metabolism and the products thereof, including any small molecules, macromolecules or chemical compounds, such as but not limited to starches, sugars, proteins, peptides, enzymes, hormones, growth factors, nucleic acid molecules, celluloses, hemicelluloses, calloses, lectins, fibers, pigments such as anthocyanins, vitamins, minerals, micronutrients, or macronutrients, that are produced by plants.

The term "plant physiology" or the term "plant physiological characteristic" or similar term will, when used herein, be understood to include the functional processes of a plant, including developmental processes such as growth, expansion and differentiation, sexual development, sexual reproduction, seed set, seed development, grain filling, asexual reproduction, cell division, dormancy, germination, light adaptation, photosynthesis, leaf expansion, fiber production, secondary growth or wood production, amongst others; responses of a plant to externally-applied factors such as metals, chemicals, hormones, growth factors, environment and environmental stress factors (*e.g.*, anoxia, hypoxia, high temperature, low temperature, dehydration, light, day length, flooding, salt, heavy metals, amongst others), including adaptive responses of plants to said externally-applied factors.

The DIM and DIMIC molecules of the present invention are useful in agriculture. The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used to modulate the protein levels or activity of a protein involved in the cell cycle, pre-mRNA processing or vesicle transport/processing.

Thus, the DIM and/or DIMIC molecules of the present invention may be used to modulate, *e.g.*, enhance, crop yields; modulate, *e.g.*, attenuate, stress, *e.g.*, heat or nutrient deprivation, tolerance; modulate tolerance to pests and diseases; modulate plant architecture; modulate plant quality traits; or modulate plant reproduction and seed development.

The DIM and/or DIMIC molecules of the present invention may also be used to modulate endoreduplication in storage cells, storage tissues, and/or storage organs of plants or parts thereof. The term "endoreduplication" includes recurrent DNA replication without consequent mitosis and cytokinesis. Preferred target storage organs and parts thereof for the modulation of endoreduplication are, for example, seeds (such as from cereals, oilseed crops), roots (such as in sugar beet), tubers (such as in potatoes) and fruits (such as in vegetables and fruit species). Increased endoreduplication in storage

organs, and parts thereof, correlates with enhanced storage capacity and, thus, with improved yield. In another embodiment of the invention, the endoreduplication of a whole plant is modulated. Grain yield in crop plants is largely a function of the amount of starch produced in the endosperm of the seed. The amount of protein produced in the endosperm is also a contributing factor to grain yield (Traas *et al.* (1998) *Current Opin. Plant Biol.* 1, 498-503). In contrast, the embryo and aleurone layers contribute little in terms of the total weight of the mature grain. By virtue of being linked to cell expansion and metabolic activity, endoreduplication is generally considered to be an important factor for increasing yield . As grain endosperm development initially includes extensive endoreduplication (Olsen *et al.* (1999) *Trends Plant Sci.* 4, 253-257), enhancing, promoting or stimulating this process is likely to result in increased grain yield. Enhancing, promoting or stimulating cell division during seed development as described *supra* is an alternative way to increase grain yield. In another aspect, the present invention also features a method for the production of SiO_2 from the peels or husks of larger rice seeds. Methods for extraction and/or production of pure SiO_2 from rice seed peels or husks are known in the art (e.g. Gorthy and Pudukottah 1999) and units for production of SiO_2 from rice seed peels are being set up (visit e.g. http://bisnis.doc.gov/bisnis/leads/990604sp.htm). SiO_2 has many applications including electronics, perfume industry and pharmacology and silicone production.

Ectopic expression, preferably downregulation of expression, of DIM and/or DIMICs may also confer enhanced resistance to pathogens causing neoplastic plant growth, such as plant pathogenic bacteria including *Agrobacterium tumefaciens, Rhodococcus fascians, Pseudomonas savastanoi, Xanthomonas campestris pv citri* and *Erwinia herbicola*, plant pathogenic fungi including *Plasmodiophora brassicae, Crinipellis perniciosa, Pucciniastrum geoppertianum, Taphrina wiesneri, Ustilaga maydis, Exobasidium vaccinii, E. camelliae, Entorrhiza casparyana* and *Apiosporina morbosum* and plant pathogenic gall-inducing insects including the midge *Mayetiola poae.*

Ectopic expression, preferably downregulation of expression, of a DIM and/or DIMIC molecule may also confer enhanced resistance or tolerance against pathogens which rely on endoreduplication events in the infected host cells to survive. The ectopic expression, preferably downregulation of expression, of a DIM and/or DIMIC molecule is expected to inhibit endoreduplication events. Pathogens relying on host cell endoreduplication to, for example, establish a feeding structure, include nematodes such as *Heterodera* species and *Meloidogyne* species.

B. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to DIMIC proteins, have a stimulatory or inhibitory effect on, for example, DIMIC expression or DIMIC activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a DIMIC substrate.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a DIMIC protein or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a DIMIC protein or polypeptide or biologically active portion thereof, *e.g.*, modulate the ability of DIMIC to interact with its cognate ligand. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a DIMIC target molecule (*e.g.*, a plant cyclin dependent kinase) with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the DIMIC target molecule. Determining the ability of the test compound to modulate the activity of a DIMIC target molecule can be accomplished, for example, by determining the ability of the DIMIC protein to bind to or interact with the DIMIC target molecule, or by determining the ability of the target molecule, *e.g.*, the plant cyclin dependent kinase, to phosphorylate a protein.

The ability of the target molecule, *e.g.*, the plant cyclin dependent kinase, to phosphorylate a protein can be determined by, for example, an *in vitro* kinase assay. Briefly, a protein can be incubated with the target molecule, *e.g.*, the plant cyclin dependent kinase, and radioactive ATP, *e.g.*, [□-^{32}P] ATP, in a buffer containing $MgCl_2$ and $MnCl_2$, *e.g.*, 10 mM $MgCl_2$ and 5 mM $MnCl_2$. Following the incubation, the immunoprecipitated protein can be separated by SDS-polyacrylamide gel electrophoresis under reducing conditions, transferred to a membrane, *e.g.*, a PVDF membrane, and autoradiographed. The appearance of detectable bands on the autoradiograph indicates that the protein has been phosphorylated. Phosphoaminoacid analysis of the phosphorylated substrate can also be performed in order to determine which residues on the protein are phosphorylated. Briefly, the radiophosphorylated protein band can be excised from the SDS gel and subjected to partial acid hydrolysis. The products can then be separated by one-dimensional electrophoresis and analyzed on, for example, a phosphoimager and compared to ninhydrin-stained phosphoaminoacid standards.

Determining the ability of the DIMIC protein to bind to or interact with a DIMIC target molecule can be accomplished by determining direct binding. Determining the ability of the DIMIC protein to bind to or interact with a DIMIC target molecule can be accomplished, for example, by coupling the DIMIC protein with a radioisotope or enzymatic label such that binding of the DIMIC protein to a DIMIC target molecule can be determined by detecting the labeled DIMIC protein in a complex. For example, DIMIC molecules, *e.g.*, DIMIC proteins, can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, DIMIC molecules can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a compound to modulate the interaction between DIMIC and its target molecule, without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of DIMIC with its target molecule without the labeling of either DIMIC or the target molecule. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between compound and receptor.

In a preferred embodiment, determining the ability of the DIMIC protein to bind to or interact with a DIMIC target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*e.g.*, intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, chloramphenicol acetyl transferase), or detecting a target-regulated cellular response.

In yet another embodiment, an assay of the present invention is a cell-free assay in which a DIMIC protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the DIMIC protein or biologically active portion thereof is determined. Binding of the test compound to the DIMIC protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the DIMIC protein or biologically active portion thereof with a known compound which binds DIMIC to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a DIMIC protein, wherein determining the ability of the test compound to interact with a DIMIC protein comprises determining the ability of the test compound to preferentially bind to DIMIC or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a DIMIC protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the DIMIC protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a DIMIC protein can be accomplished, for example,

by determining the ability of the DIMIC protein to bind to a DIMIC target molecule by one of the methods described above for determining direct binding. Determining the ability of the DIMIC protein to bind to a DIMIC target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a DIMIC protein can be accomplished by determining the ability of the DIMIC protein to further modulate the activity of a DIMIC target molecule (*e.g.*, a DIMIC mediated signal transduction pathway component). For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a DIMIC protein or biologically active portion thereof with a known compound which binds the DIMIC protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the DIMIC protein, wherein determining the ability of the test compound to interact with the DIMIC protein comprises determining the ability of the DIMIC protein to preferentially bind to or modulate the activity of a DIMIC target molecule.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of proteins (*e.g.*, DIMIC proteins or biologically active portions thereof). In the case of cell-free assays in which a membrane-bound form a protein is used it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either DIMIC or its target molecule to facilitate

separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a DIMIC protein, or interaction of a DIMIC protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ DIMIC fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or DIMIC protein, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of DIMIC binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a DIMIC protein or a DIMIC target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated DIMIC protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with DIMIC protein or target molecules but which do not interfere with binding of the DIMIC protein to its target molecule can be derivatized to the wells of the plate, and unbound target or DIMIC protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the DIMIC protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the DIMIC protein or target molecule.

In another embodiment, modulators of DIMIC expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of DIMIC mRNA or protein in the cell is determined. The level of expression of DIMIC

mRNA or protein in the presence of the candidate compound is compared to the level of expression of DIMIC mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of DIMIC expression based on this comparison. For example, when expression of DIMIC mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of DIMIC mRNA or protein expression. Alternatively, when expression of DIMIC mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of DIMIC mRNA or protein expression. The level of DIMIC mRNA or protein expression in the cells can be determined by methods described herein for detecting DIMIC mRNA or protein.

In yet another aspect of the invention, the DIMIC proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with DIMIC ("DIMIC-binding proteins" or "DIMIC-bp") and are involved in DIMIC activity. Such DIMIC-binding proteins are also likely to be involved in the propagation of signals by the DIMIC proteins or DIMIC targets as, for example, downstream elements of a DIMIC-mediated signaling pathway. Alternatively, such DIMIC-binding proteins are likely to be DIMIC inhibitors.

Alternatively, a mammalian two-hybrid system can be used which includes e.g. a chimeric green fluorescent protein encoding reporter gene (Shioda et al. 2000, Proc. Natl. Acad. Sci. USA 97, 5520-5224). Yet another alternative consists of a bacterial two-hybrid system using e.g. *HIS* as reporter gene (Joung et al. 2000, Proc. Natl. Acad. Sci. USA 97, 7382-7387).

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a DIMIC protein is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a DIMIC-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close

proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the DIMIC protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate plant or animal model. For example, an agent identified as described herein (*e.g.*, a DIMIC modulating agent, an antisense DIMIC nucleic acid molecule, a DIMIC-specific antibody, or a DIMIC-binding partner) can be used in a plant or animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in a plant or animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for the agricultutal and therapeutic uses described herein.

C. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; identify an individual from a minute biological sample (tissue typing); and aid in forensic identification of a biological sample. Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the DIMIC nucleotide sequences, described herein, can be used to map the location of the DIMIC genes on a chromosome. The mapping of the DIMIC sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, DIMIC genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the DIMIC nucleotide sequences. Computer analysis of the DIMIC sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of cell hybrids containing individual plant or human

chromosomes. Only those hybrids containing the plant or human gene corresponding to the DIMIC sequences will yield an amplified fragment.

Other mapping strategies which can similarly be used to map a DIMIC sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature*, 325:783-787.

Moreover, differences in the DNA sequences between plants affected and unaffected with a disease associated with the DIMIC gene, can be determined. If a mutation is observed in some or all of the affected plants but not in any unaffected plants,

then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected plants generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several plants can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

D. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining DIMIC protein and/or nucleic acid expression as well as DIMIC activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant DIMIC expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with DIMIC protein, nucleic acid expression or activity. For example, mutations in a DIMIC gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby phophylactically treat an individual prior to the onset of a disorder characterized by or associated with DIMIC protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of DIMIC in clinical trials.

E. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant DIMIC expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug

response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the DIMIC molecules of the present invention or DIMIC modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Figures and the Sequence Listing are incorporated herein by reference.

EXAMPLES

EXAMPLE 1: IDENTIFICATION OF PLANT DIM1 PROTEINS

The amino acid sequence of the *Schizosaccharomyces pombe* dim1p protein (GenBank accession number AF001214; protein ID AAC49744.1) was used to perform a blastp search in the publicly available GenBank database. The nucleotide sequence corresponding to the protein hit found in *Arabidopsis thaliana* (GenBank accession number AL392174; protein ID CAC08329.1) was subsequently used to perform an additional blastn search in the publicly available GenBank database. A number of plant homologues of the *S. pombe* dim1p protein were identified as summarized in Table III. The amino acid sequences of these proteins are aligned in Figure 1. For comparison purposes, the *S. pombe* and *Drosophila melanogaster* DIM1 sequences are also included in Table III and Figure 1.

The DIM1 homologues invariably carry at their amino-termini the motif characteristic for the FKBP-type peptidyl-prolyl *cis-trans* isomerases: [L/I/V/M/C] X [Y/F] X [G/V/L] X_{1-2} [L/F/T] X_2 G X_3 [D/E] [S/T/A/E/Q/K] [S/T/A/N] (SEQ ID NO 54), with 'X' being any amino acid residue, 'Xn' being a stretch of 'n' random amino acid residues and, *e.g.*, '[Y/F]' meaning either a tyrosine or phenylalanine residue occurring at that position.

The nearly perfect conservation of the identified plant DIM1 proteins implies that the DIMIC proteins of the present invention interact not only with the Arabidopsis DIM1, but almost certainly also with the DIM1 proteins of all other plants.

Table III. Overview of the identified plant, fission yeast and Drosophila DIM homologs. Indicated are the GenBank accesion numbers of DNA and protein as well as the SEQ ID NO defining said DNA and protein sequences.

Organism	GenBank accession number	GenBank protein ID number	SEQ ID NO: (DNA)	SEQ ID NO: (protein)
Arabidopsis thaliana	AL392174	CAC08329.1	1	18
Glycine max	AW203293	/	2	19
Medicago truncatula	AJ388790	/	3	20
Lycopersicon esculentum	AW650896	/	4	21
Gossypium arboreum	BE055159	/	5	22
Lotus japonica	AV417895	/	6 (partial)	23 (partial)
Zea mays	AI795355	/	7	24
Oryza sativa	AU094120; AU063980	/	8	25
Hybrid aspen	AI165558; AI162529	/ /	9	26
Pinus taeda	AW289752	/	10	27
Triticum aestivum	BE426097	/	11	28
Hordeum vulgare	BE422314	/	12	29
Thellungiella salsuginea	BE727006	/	13 (partial)	30 (partial)
Cryptomeria japonica	AU083780	/	14 (partial)	31 (partial)
Mesembryanthemum crystallinum	BE577228	/	15 (partial)	32 (partial)
Schizosaccharomyces pombe	AF001214	AAC49744.1 P87215	16	33
Drosophila melanogaster	AE003576	AAF51017	17	34

5

EXAMPLE 2: IDENTIFICATION OF DIM AND DIM-INTERACTING CLONES

A two-hybrid screening was performed using as a bait a fusion between the GAL4 DNA-binding domain and *DIM1At*. Vectors and strains used were provided with the Matchmaker Two-Hybrid System (Clontech, Palo Alto, CA). The bait was constructed by inserting the *DIM1At* PCR fragment into the pGBT9 vector. The PCR fragment was created from the cDNA using a sense primer incorporating an *Eco*RI restriction enzyme site (5'-GGGAATTCATGTCGTATCTTCTTCCACATCTGC-3', *Eco*RI-site underlined; SEQ ID NO 71) and an antisense primer incorporating a *Bam*HI restricition enzyme site (5'-GGGGATCCAAATTTCATTTCACTGAATCATGTTCG-3', *Bam*HI-site underlined; SEQ ID NO 72). The PCR fragment was cut with *Eco*RI and *Bam*HI and cloned into the *Eco*RI and *Bam*HI sites of pGBT9, resulting in the plasmid pGBTDIM1. The GAL4 activation domain cDNA fusion library was constructed using mRNA of *Arabidopsis thaliana* cell suspension cultures harvested at various growing stages: early exponential, exponential, early stationary, and stationary phase.

For the two-hybrid screening a 1-liter culture of the *Saccharomyces cerevisiae* strain PJ69-4A containing the pGBTDIM1 plasmid was transformed with 120 µg DNA of the library, and 1000 µg herring testes carrier DNA using the lithium acetate method (Gietz *et al.*, 1992). To estimate the number of independent cotransformants, 1/1000 of the transformation mix was plated on Leu- and Trp- medium. The rest of the transformation mix was plated on the medium to select for histidine prototrophy (Trp-, Leu-, His-, +10mM 3AT). After 5 days of growth at 30°C, colonies larger than 2 mm were streaked on adenine lacking medium.

A total of 10^6 independent cotransformants were screened for their ability to grow on adenine free medium. A 5-day incubation at 30°C yielded 28 colonies growing on adenine free medium.

Of the Ade^+ colonies the activation domain plasmids were isolated as described (Hoffman and Winston, 1987). The pGAD10 inserts were PCR amplified using the primers 5'-ATACCACTACAATGGATG-3' (SEQ ID NO 73) and 5'-AGTTGAAGTGAACTTGCGGG-3' (SEQ ID NO 74). PCR fragments were digested with *Alu*I and fractionized on a 2% agarose gel. Plasmid DNA, the inserts of which gave rise to different restriction patterns, was electroporated into *Escherichia coli* XL1-Blue, and the DNA sequence of the inserts was determined. Extracted DNA was also used to retransform PJ69-4A to test the specificity of the interaction.

After the sequential selection rounds four true positive clones were identified and are defined herein as "DIM-Interacting Clones" or "DIMICs."

EXAMPLE 3: CHARACTERIZATION OF DIMIC5

The partial nucleotide sequence of the DIMIC5 cDNA was determined (SEQ ID NO 35) and a full-length sequence reconstituted based on GenBank entries AC004261 (ORF corresponding to protein ID AAD12009), T02117 (hypothetical protein T3K9.21 which is identical to AAD12009) and T02116. None of these entries, however, fully matches with DIMIC5. The protein AAD12009/T02117 extends the partial DIMIC5 protein derived from the partial DIMIC5 cDNA by 88 amino acids. The further 237 amino acids of AAD12009/T02117 are identical to the 237 N-terminal amino acids of the partial DIMIC5 protein. The partial DIMIC5 protein has, however, relatively to AAD12009/T02117 an additional C-terminal region consisting of 138 amino acids. The C-terminal region is partially covered by T02116. However, T02116 comprises an internal stretch of 31 amino acids not present in the partial DIMIC5 protein. Based on these observations, the full-length DIMIC5 amino acid and nucleotide sequences were reconstituted by adding the 88 N-terminal amino acids of AAD12009 and the corresponding nucleotides of AC004261, respectively. The full-length DIMIC5 protein sequence is set forth in Figure 2 and in SEQ ID NO 49. The region of the Arabidopsis genome covering the DIMIC5 open reading frame is shown in Figure 3, defined in SEQ ID NO 37 and corresponds to nucleotides 17241 to 20717 of AC004261. Also indicated in Figure 3 is the full-length DIMIC5 cDNA sequence (SEQ ID NO 36) interrupted by the intron sequences. Both the DIMIC5 protein and cDNA thus are novel molecules as such not present in the GenBank database.

A closer analysis of the DIMIC5 protein revealed the presence of two WW-domains (consensus sequence [W X_{22} W X_2 P] (SEQ ID NO 55) with X_n being a stretch of n random amino acid residues) organized in tandem followed by a C-terminal domain comprising a non-classical C_2-domain (SEQ ID NO 56). The C-terminal domain is highly homologous to the C-terminal domain of human PQBP-1 (polyglutamine tract-binding protein) proteins (Waragai *et al.* (1999) *Human Mol Genet* 8, 977-987; GenBank accession number AJ242829) and mouse PQBP-1 (GenBank accession number NM019478) both of which contain a C_2-domain. WW domains represent small and compact globular structures that interact with proline-rich ligands (Bedford *et al.* (1997) *EMBO J* 16, 2376-2383; Chan *et al.* (1996) *EMBO J* 15, 1045-1054; Einbond and Sudol

(1996) *FEBS Lett* 384, 1-8). Most proteins containing C_2-domains are functional in signal transduction or membrane trafficking. Phospholipid binding to many C_2-domains is regulated by Ca^{2+} and C_2-domain proteins are, therefore, implicated in Ca^{2+}-dependent phospholipid signalling (Rizo and Südhof (1998) *J Biol Chem* 273, 15879-15882).

An additional intramolecularly repeated motif, termed DIMIC5 internal repeat domain1, was furthermore discerned by dot plot analysis of the DIMIC5 protein sequence (Omiga 2.0 software; scoring matrix: Blosum 62; stringency: 60%; window: 8; hash size: 2). This motif is indicated in Figure 2 and consists of the amino acid sequence GGWXVGL (SEQ ID NO 57) wherein X is any amino acid.

Using the PESTFIND software (downloadable from http://www.ebi.ac.uk), a potential PEST sequence with a PEST-FIND score 14 was identified in the DIMIC5 protein with sequence 'HAEDDELDPMDPSSYSDAPR' (residues 373-392 of SEQ ID NO 49).

EXAMPLE 4: CHARACTERIZATION OF DIMIC7=DIMIC40

The partial nucleotide sequence of the DIMIC7 cDNA was determined (SEQ ID NO 38) and shown to be identical to a second DIMIC clone, DIMIC40, that was identified independently of DIMIC7. The full-length DIMIC7 sequences were reconstituted based on GenBank entry AC008148 (ORF corresponding to protein ID AAD55502; unknown protein). The region of the Arabidopsis genome covering the DIMIC7=DIMIC40 ORF is shown in Figure 4 and defined by SEQ ID NO 40. This region corresponds to nucleotides 100439-106312 of AC008148. Also indicated in Figure 4 is the full-length DIMIC7=DIMIC40 cDNA sequence (SEQ ID NO 39) interrupted by intron sequences. The full-length DIMIC7=DIMIC40 protein is shown in Figure 5 (and defined in SEQ ID NO 50) with an indication of the partial amino acid sequence as derived from the partial DIMIC7=DIMIC40 cDNA.

The full-length DIMIC7=DIMIC40 protein displays 41% identity / 51% similarity to the *A. thaliana* protein AtFAB1 (GenBank accession number AL035525, protein ID CAB36798). The AtFAB1 protein is aligned with DIMIC7 in Figure 5. AtFAB1 is known in the art as a type III phosphatidylinositol 3-phosphate 5-kinase (PtdIns3*P* 5-kinase; Cooke *et al.* (1998) *Current Biol* 8, 1219-1222; McEwen *et al.* (1999) *J Biol Chem* 274, 33905-33912). Characteristic for FAB1 proteins is the conservation of three domains: (i) the FYVE zinc-finger domain (consensus sequence [R/K][R/K]HHCR); (ii) a CCT-homology domain; and (iii) the catalytic domain (McEwen *et al.* (1999) *J Biol Chem* 274, 33905-

33912). The FYVE-domain specifically binds to PtdIns3*P in vitro* (Odorizzi *et al.* (2000) *TIBS* 25, 229-235). CCT or 'Chaperonine Containg TCP-1' is a cytosolic hetero-oligomeric chaperone acting on a limited number of substrates. The CCTδ , CCTβ , and CCTε subunits have, for example, been shown to interact with α-actin (Llorca *et al.* (1999) *Nature* 402, 693-696) whereas TCP-1 activity is required for growth of microtubules off the centrosome (Brown *et al.* (1996) *J Biol Chem* 271, 824-832). The FYVE-domain is present in AtFAB1 (see Figure 5) but is lacking in DIMIC7=DIMIC40. The CCT-homology domain is conserved in DIMIC7=DIMIC40 as is illustrated in Figure 5 where the corresponding regions of the AtFAB1, the *S. cerevisiae* Fab1p (GenBank accession number P34756) and of the mouse CCTδ (GenBank accession number Z31554) proteins are aligned with DIMIC7=DIMIC40. The C-terminal catalytic domain characteristic of the FAB1 kinases is also present and conserved in DIMIC7=DIMIC40 as is obvious from Figure 5 (alignment of AtFAB1 and yeast Fab1p catalytic domains with the corresponding region of DIMIC7). Residues invariant in the catalytic domains of these proteins are indicated by an asterisk in Figure 5 and these include the residues invariant in all FAB1 kinases known to date: (i) K2059; (ii) D2196; and (iii) D2216 (numbering of all three residues relative to the yeast Fab1p protein. The FAB1 activation loop with consensus sequence T[F/Y]T[W/L]DKKLE[S/T/M]WVKXXG[I/L][V/L]G (SEQ ID NO 58) is, with the exception of the first threonine residue (which is a glutamine in DIMIC7), completely conserved in DIMIC7 (see Figure 5). This motif may be involved in defining PtdIns3*P* as the substrate for 5-phosphorylation (McEwen *et al.* (1999) *J Biol Chem* 274, 33905-33912).

Five different intramolecularly repeated motifs, termed DIMIC7 internal repeat domains 1-5, were furthermore discerned by dot plot analysis of the DIMIC7 protein sequence (Omiga 2.0 software; scoring matrix: Blosum62; stringency: 60%; window: 8; hash size: 2). These motifs are depicted in Figure 6 and are:

Motif DIMIC7/1: PLGR[F/W/Y][M/I/L/V] (SEQ ID NO 59);

Motif DIMIC7/2: EXXG[R/K/H]IW (SEQ ID NO 60);

Motif DIMIC7/3: DLXXPT[M/I/L/V] (SEQ ID NO 61);

Motif DIMIC7/4: DDXXSXYF (SEQ ID NO 62);and

Motif DIMIC7/5: TEXSDXLN (SEQ ID NO 63); with X being any amino acid and, *e.g.*, [D/E] being either an aspartate or glutamate residue at that position.

Phosphoinositides are generally known as key regulators of vesicle-mediated protein trafficking. PtdIns(3,5)P_2, in particular, was shown to be required for normal vacuolar morphology and function. In the absence of active Fab1p, yeast cells

accumulate large vacuoles with a reduced hydrolytic activity as the result of poor acidification. The number of multivesicular bodies (MVBs) is also strongly reduced. MVBs target membrane proteins for vacuolar degradation by fusing with the vacuole. In the absence of functional Fab1p, however, the membrane proteins are delivered to the vacuolar outer membrane instead of into the vacuolar lumen. This process could explain the increase in vacuole size (Odorizzi *et al.* (1998) *Cell* 95, 847-858; Odorizzi *et al.* (2000) *TIBS* 25, 229-235). The vacuolar surface area increases up to 2.5-fold and is probably the cause of inappropriate nuclear segregation during mitosis (Gary *et al.* (1998) *J Cell Biol* 143, 65-79).

Interestingly, Fab1p is also likely to be involved, via PtdIns(3,5)P_2, in the acute osmoprotective response which is accompanied by or requires enhanced PtdIns(3,5)P_2-dependent vesicle trafficking. PtdIns(3,5)P_2 is present in yeast as well as in mammalian and plant cells. Hyperosmotically stressed yeast and mammalian cells display a rapid increase and decrease, respectively, in the levels of PtdIns(3,5)P_2 whereas hypo-osmotically stressed mammalian cells rapidly accumulate high levels of PtdIns(3,5)P_2 (Dove *et al.* (1997) *Nature* 390, 187-192). Fab1p might furthermore be involved in retrograde transport from the vacuole via PtdIns(3,5)P_2 synthesis and the recruitment of PtdIns(3,5)P_2-binding proteins (Cooke *et al.* (1998) *Current Biol* 8, 1219-1222).

FAB1 homologues are found in plants other than *Arabidopsis thaliana*, *e.g.*, in rice (GenBank accession number C28212) and date palm (Corniquel and Mercier (1997) *Int J Plant Sci* 158, 152-156).

Using the PESTFIND software (downloadable from http://www.ebi.ac.uk), eight potential PEST sequences were identified in the DIMIC7 protein. These PEST sequences are given in Table IV.

Table IV. Overview of the potential PEST sequences identified in the DIMIC7 protein.

PEST-sequence number	Amino acid residues in SEQ ID NO 50	amino acid sequence	PEST-FIND score
1	103-126	RDSPDPPSSLATESESLASSLEIR (SEQ ID NO 98)	12.4
2	151-175	KQLLSPSSDNYQDSSDIESGSVSAR (SEQ ID NO 99)	2.9
3	262-282	RIWYPPPPEDENDDAESNYFH (SEQ ID NO 100)	4.8
4	661-702	HEVCESLCEDFDPTQIFPPSSEVETEQSDTLNGD FANNLVTR (SEQ ID NO 101)	4.9
5	713-739	HEPTLCLSSEIPETPTQQPSGEEDNGR (SEQ ID NO 102)	13
6	996-1011	KSSLLEPEQSEACDLH (SEQ ID NO 103)	0.2
7	1127-1142	KDPENIPSPGTSLSER (SEQ ID NO 104)	10
8	1320-1331	RESEPSAFSTWR (SEQ ID NO 105)	1

EXAMPLE 5: CHARACTERIZATION OF DIMIC26

The partial nucleotide of the DIMIC26 cDNA was determined (SEQ ID NO 41) and a full-length sequence was reconstituted based on GenBank entry AB023039 (ORF corresponding to protein ID BAA96996). The region of the Arabidopsis genome covering the DIMIC26 ORF is shown in Figure 7 and defined in SEQ ID NO 43. This region corresponds to the inverse complement of nucleotides 19634-21435 of AB023039. Also indicated in Figure 7 is the full-length DIMIC26 cDNA sequence (SEQ ID NO 42) interrupted by one intron sequence. The full-length DIMIC26 protein is shown in Figure 8 and is defined in SEQ ID NO 51 with an indication of the partial amino acid sequence as derived from the partial DIMIC26 cDNA. The function of DIMIC26 is currently not known. The presence, however, of DIMIC26 regions with weak homology to parts of the human CENP-E (centrosome protein E; GenBank accession number NM001813) and NMMHC-B (nonmuscle type B myosin heavy chain; GenBank accession number P35580) point at a role of DIMIC26 in cell cycle processes such as chromosome movement and/or spindle elongation and/or cytokinesis. Such a role is in line with the reported function of DIM1 in chromosome segregation and the localization of DIM1 to the mitotic spindle (*supra*). The

alignment of the homologous regions of DIMIC26 and CENP-E (37% identity, 53% similiarity) is given in Figure 9, and the alignment of the homologous regions of DIMIC26 and NMMHC-B (29% identity, 41% similarity) is given in Figure 10.

Six different intramolecularly repeated motifs, termed DIMIC26 internal repeat domains 1-6, were furthermore discerned by dot plot analysis of the DIMIC26 protein sequence (Omiga 2.0 software; scoring matrix: Blosum62; stringency: 60%; window: 8; hash size: 2). These motifs are depicted in Figure 11 and are:

Motif DIMIC26/1: CXCXIC (SEQ ID NO 64);

Motif DIMIC26/2: ACNRXXE[M/I/L/V][M/I/L/V](SEQ ID NO 65);

Motif DIMIC26/3: QXSGGG (SEQ ID NO 66);

Motif DIMIC26/4: [M/I/L/V]DX[M/I/L/V]KXGL (SEQ ID NO 67);

Motif DIMIC26/5: SEXXAEKQ(SEQ ID NO 68); and

Motif DIMIC26/6: RLXXAEA[D/E](SEQ ID NO 69); with X being any amino acid and, *e.g.*, [D/E] being either an aspartate or glutamate residue at that position.

Further present in DIMIC26 are the '[M/I/L/V][R/K/H]' amino acid pair and the '[R/K/H][M/I/L/V]' amino acid pair which are repeated multiple times. Note that both pairs can overlap.

Using the PESTFIND software (downloadable from http://www.ebi.ac.uk), one potential PEST sequence was identified in the DIMIC26 protein. Said PEST sequences comprises amino acid residues 41-63 of SEQ ID NO 51, has a PEST-FIND score of 6.3 and comprises the sequence 'RESPAESASSQETWPLGDTVAGK' (SEQ ID NO 107).

Within the DIMIC26 protein sequence (SEQ ID Nr. 51), a PHD finger motif (InterPro Accession number IPR001965 and IPR001841) was found with the software program InterPro (<http://www.ebi.ac.uk/interpro>). This domain is positioned from amino acid 218 till 291 with conserved cystein residues in between: (SEQ ID NO 111 : NRKGFCNLCMCTICNKFDFSVNTCRWIGCDLCSHWTHTDCAIRDGQITTGSSAKNNTSG PGEIVFKCRACNRT).This domain has implications for chromatin-mediated transcriptional regulation and therefor constitutes an important active domain of the DIMID26 protein.

EXAMPLE 6: CHARACTERIZATION OF DIMIC70

The partial nucleotide sequence of the DIMIC70 cDNA was determined (SEQ ID NO 44) and a full-length sequence was reconstituted based on GenBank entry AC007583 (ORF corresponding to protein ID AAF75085). Thus, the partial DIMIC70 protein

sequence as derived from the partial DIMIC70 cDNA was N-terminally extended with the N-terminal 126 amino acids of AAF75085. The region of the Arabidopsis genome covering the DIMIC70 ORF is shown in Figure 12 and defined by SEQ ID NO 47. This region corresponds to the inverse complement of nucleotides 64105-65587 of AC007583. The predicted protein AAF75085 is, however, lacking a stretch of 37 amino acids that is present in the partial DIMIC70 sequence, most likely as the result of wrong intron-exon prediction in the ORF corresponding to AAF75085. In addition, the DNA region encoding the 37-amino acid stretch of DIMIC70 is shorter with 3 nucleotides compared to the corresponding region of the genomic DNA fragment (nucleotides "tga" at positions 64486-64484 in SEQ ID NO 47). Thus, probably as a result of allelic variation, it can be concluded that two DIMIC70 gene, cDNA and protein forms exist which are referred to as DIMIC70A (full-length protein shown in Figure 13B and defined by SEQ ID NO 52; the full-length DIMIC70A cDNA sequence is given in Figure 13A and defined by SEQ ID NO 45; the gene defined by SEQ ID NO 47) and DIMIC70B (full-length protein shown in Figure 14B and defined by SEQ ID NO 53; the full-length DIMIC70B cDNA sequence is given in Figure 14A and defined by SEQ ID NO 46; the gene defined by SEQ ID NO 48). Compared to the DIMIC70A protein the DIMIC70B protein contains an additional amino acid "D" or aspartate at position 203 of DIMIC70B (SEQ ID NO 53; Figure 14B).

Both the DIMIC70A and DIMIC70B proteins and cDNAs thus are novel molecules as such not present in the GenBank database.

A third DIMIC70 molecule is represented in figure 16 (A: nucleic acid sequence; B; amino acid sequence) and represents a third variant cDNA sequence. The DIMIC 70C protein (SEQ ID NO 95) is shorter than the DIMIC70B protein and differs in the first 17 amino acids.

The (thioredoxin –like protein) motif described is present in the DIMIC70C protein sequence (SEQ ID NO 95);

The C-terminal part of DIMIC70 displays a 34% identity and a 45% similarity to the PRODOM family PD012637 consensus sequence which comprises the redox-active center of thioredoxins and thioredoxin-like proteins. The consensus thioredoxin-like domain 'CXXC' (SEQ ID NO 70) wherein 'X' is any amino acid (Wang and Chang (1999) *EMBO J* 18, 5972-5982) together with the alignment of PD012637 with the homologous DIMIC70 region is indicated in Figure 15. Thus, unlike DIM1, DIMIC70 does contain a candidate thioredoxin active center and is, thus, a putative active thioredoxin. Without being intenting to be bound by any theory or mode of action, it is postulated that DIMIC70 could be involved in the process leading to the proper protein folding of DIM1.

Using the PESTFIND software (downloadable from http://www.ebi.ac.uk), one potential PEST sequence was identified in the DIMIC70A protein as well as in the DIMIC70B and DIMIC70C protein. These PEST sequences comprise amino acid residues 129-156 of SEQ ID NO 52 and of SEQ ID NO 53, have a PEST-FIND score of 1.0 and comprise the sequence 'KSLSQENLVELSDENDDLCPVECVTEFK' (SEQ ID NO 110).

EXAMPLE 7: EXPRESSION OF RECOMBINANT DIMIC PROTEINS IN TRANSGENIC PLANTS

In this example, the DIMIC molecules of the present invention were expressed in a ^{35}S expression vector in transgenic plants. The DIMIC molecules of this invention were cloned using standard cloning procedures between the *CaMV ^{35}S* promoter and the NOS 3' untranslated region in the *Nco*I and *Bam*HI sites of PH^{35}S (Hemerly *et al,* EMBO J.14 (1995), 3925-3936), resulting in the PH^{35}SDIMIC vector. This construct was cloned in the binary vector PSV4 (Herouart *et al.*, Plant Physiol. 104 (1994), 873-880) and in Agrobacterium tumefaciens. The constructs were introduced in Nicotiana tabacum cv. Petit havana (SR1) plants by the leaf disk protocol (Horsh, Science 227 (1985), 1229-1231).

For tissue-specific expression, the DIMIC gene is expressed under control of the minimal ^{35}S promoter containing UAS elements. These UAS elements are sites for transcriptional activation by the GAL4-VP16 fusion protein. The GAL4-VP16 fusion protein in turn is expressed under control of a tissue-specific promoter. The UAS-DIMIC construct and the GAL4-VP16 construct are combined by co-transformation of both constructs, subsequent transformation of single constructs or by sexual cross of lines that contain the single constructs. The advantage of this two-component system is that a wide array of tissue-specific expression patterns can be generated for a specific transgene, by simply crossing selected parent lines expressing the UAS-DIMIC construct with various tissue-specific GAL4-VP16 lines. A tissue-specific promoter/DIMIC combination that gives a desired phenotype can subsequently be recloned in a single expression vector, to avoid stacking of transgene constructs in commercial lines.

Primary transformants were selfed and characterized by Northern and Western blotting using 3 week old plantlets. Expression levels were compared with those of non-transformed (control) plants.

EXAMPLE 8: DOWNREGULATION OF TARGET DIMIC GENES IN TRANSGENIC PLANTS

Plant genes can be specifically downregulated by antisense and co-suppression technologies. These technologies are based on the synthesis of antisense transcripts, complementary to the mRNA of a given DIMIC gene. There are several methods described in literature, that increase the efficiency of this downregulation, for example to express the sense strand with introduced inverted repeats, rather than the antisense strand. The constructs for downregulation of target genes are made similarly as those for expression of recombinant proteins, *i.e.*, they are fused to promoter sequences and transcription termination sequences. Promoters used for this purpose are constitutive promoters as well as tissue-specific promoters.

EXAMPLE 9: *AGROBACTERIUM*-MEDIATED RICE TRANSFORMATION

Mature dry seeds of the rice japonica cultivars Nipponbare or Taipei 309 are dehusked, sterilised and germinated on a medium containing 2,4-D (2,4-dichlorophenoxyacetic acid). After incubation in the dark for four weeks, embryogenic, scutellum-derived calli are excised and propagated on the same medium. Selected embryogenic callus is then co-cultivated with *Agrobacterium*. Widely used *Agrobacterium* strains such as LBA4404 or C58 harbouring binary T-DNA vectors can be used. The *hpt* gene in combination with hygromycin is suitable as a selectable marker system but other systems can be used. Co-cultivated callus is grown on 2,4-D-containing medium for 4 to 5 weeks in the dark in the presence of a suitable concentration of the selective agent. During this period, rapidly growing resistant callus islands develop. After transfer of this material to a medium with a reduced concentration of 2,4-D and incubation in the light, the embryogenic potential is released and shoots develop in the next four to five weeks. Shoots are excised from the callus and incubated for one week on an auxin-containing medium from which they can be transferred to the soil. Hardened shoots are grown under high humidity and short days in a phytotron. Seeds can be harvested three to five months after transplanting. The method yields single locus transformants at a rate of over 50 % (Aldemita and Hodges (1996) Planta 199, 612-617 ; Chan *et al.* (1993) *Plant Mol Biol* 22, 491-506, Hiei *et al.* (1994), *Plant J* 6, 271-282).

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

CLAIMS

1. An isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule comprising the nucleotide sequence as given in any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof,

(b) a nucleic acid molecule comprising the RNA sequence corresponding to any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof,

(c) a nucleic acid molecule specifically hybridizing with the nucleotide sequence as defined in (a) or (b),

(d) a nucleic acid molecule which is at least 60% identical to the nucleotide sequence as given in any of SEQ ID NOs 36, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 or 94, or the complement thereof,

(e) a nucleic acid molecule encoding a protein comprising an amino acid sequence as given in any of SEQ ID NOs 49 to 53 or 95,

(f) a nucleic acid molecule encoding a protein comprising at least one of the amino acid sequences represented in SEQ ID NOs 55, 56 or 96,

(g) a nucleic acid molecule encoding a protein comprising an amino acid sequence which is at least 42 % identical to the amino acid sequence as given in SEQ ID NO 50,

(h) a nucleic acid molecule encoding a protein comprising at least one of the amino acid sequences represented in SEQ ID NOs 59 to 63, or 97 to 105,

(i) a nucleic acid molecule encoding a protein comprising at least one of the amino acid sequences represented in SEQ ID NOs 64 to 69, 106, 107 or 111,

(j) a nucleic acid molecule encoding a protein comprising at least one of the amino acid sequences represented in any of SEQ ID NOs 108, 109 or 110,

(k) a nucleic acid molecule encoding a protein comprising an amino acid sequence which is at least 50 % identical to the amino acid sequence as given in any of SEQ ID NOs 49, 50, 51, 52, 53 or 95,

(l) a nucleic acid molecule which is degenerated to a nucleic acid as defined in any of (a) to (k) as a result of the genetic code,

(m) a nucleic acid molecule which is diverging from a nucleic acid as defined in any of (a) to (k) as a result of differences in codon usage between organisms,

(n) a nucleic acid molecule which is diverging from a nucleic acid as defined in any of (a) to (k) as a result of differences between alleles, and

(o) a nucleic acid molecule as defined in any one of (a) to (n) characterized in that said nucleic acid is DNA, cDNA, genomic DNA or synthetic DNA,

characterized in that said nucleic acid molecule encodes a DIM1-interacting molecule (DIMIC molecule), or a homologue or a derivative thereof and further provided that said nucleic acid is not one of the nucleic acids as deposited under the GenBank Accession numbers AC004261, AC008148, AB023039 or AC007583.

2. An isolated nucleic acid molecule encoding an immunologically active and/or functional fragment of a DIM1-interacting molecule encoded by a nucleic acid of claim 1, or an immunologically active and/or functional fragment of a homologue or a derivative of such a DIM1-interacting molecule, provided that said nucleic acid is not one of the nucleic acids as deposited under the GenBank Accession number T3K9.20 or T3K9.21.

3. An isolated nucleic acid molecule according to claim 2 selected from the group consisting of consisting of

(a) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 49, wherein said fragment comprises at least one of the sequences as represented in any of SEQ ID NOs 55, 56, or 96,

(b) a nucleic acid molecule encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 49, wherein, said fragment comprises at least 326 contiguous amino acid residues of the amino acid sequence of SEQ ID NO 49,

(c) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 50, wherein said fragment comprises at least one of the sequences as represented in any of SEQ ID NOs 59, 60, 61, 62, 63, 97, 98, 99, 100, 101, 102, 103, 104, or 105,

(d) a nucleic acid encoding a functional fragment of polypeptide comprising the amino acid sequence of SEQ ID NO 51, wherein said fragment comprises the sequence as represented in any of SEQ ID NOs 64, 65, 66, 67, 68, 69, 106, 107 or 111,

(e) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 52, wherein said fragment comprises at least one of the sequences as represented in SEQ ID NO 108 or 110,

(f) a nucleic acid molecule encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 53, wherein said fragment comprises at least one of the sequences as represented in SEQ ID NO 109 or 110,

(g) a nucleic acid encoding a functional fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO 95, wherein said fragment comprises at least one of the sequences as represented in SEQ ID NO 109 or 110, and

(h) a nucleic acid molecule encoding a functional fragment of a polypeptide comprising the amino acid sequence of any SEQ ID NOs 52, 53 or 95, wherein the fragment comprises at least 178 contiguous amino acid residues of any of the amino acid sequences of SEQ ID NOs 52, 53 or 95.

4. An antisense nucleic acid molecule corresponding to at least one of the nucleic acids as defined in any of claims 1 to 3.

5. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1 to 4 and a nucleotide sequence encoding a heterologous polypeptide.

6. A nucleic acid molecule of at least 15 contiguous nucleotides in length specifically hybridizing with a nucleic acid as defined in any of claims 1 to 3.

7. A nucleic acid molecule of at least 15 contiguous nucleotides in length specifically amplifying a nucleic acid as defined in any of claims 1 to 3.

8. A vector comprising a nucleic acid sequence as defined in any of claims 1 to 5.

9. A vector according to claim 8 which is an expression vector wherein said nucleic acid sequence is operably linked to one or more control sequences allowing the expression of said sequence in prokaryotic and/or eukaryotic host cells.

10. A host cell comprising a nucleic acid molecule as defined in any of claims 1 to 5 or a vector according to claim 8 or 9.

11. A host cell according to claim 10, wherein the host cell is a bacterial, insect, fungal, yeast, plant or animal cell.

12. An isolated polypeptide encodable by a nucleic acid as defined in any of claims 1 to 3, or a homologue or a derivative thereof, or an immunologically active and/or functional fragment thereof.

13. A polypeptide of claim 12 having an amino acid sequence as given in any of SEQ ID NOs 49 to 53 or 95, or a homologue or a derivative thereof, or an immunologically active and/or functional fragment thereof.

14. A polypeptide of claim 12 or 13 further comprising heterologous amino acid sequences or a polypeptide encodable by a nucleic acid of claim 5.

15. A method for producing a polypeptide as defined in any of claims 12 to 14 comprising culturing a host cell of claim 10 or 11 under conditions allowing the expression of the polypeptide and recovering the produced polypeptide from the culture.

16. An antibody specifically recognizing a polypeptide of claim 12 or 13 or a specific epitope of said polypeptide.

17. A method for the production of altered plant cells, plant tissues or plants comprising the introduction of a polypeptide as defined in any of claims 12 to 14 directly into said plant cell or tissue or in an organ of said plant.

18. A method for effecting the expression of a polypeptide as defined in any of claims 12 to 14 in plant cells, tissues or plants comprising the introduction of a nucleic acid molecule as defined in any of claims to 1 to 3 or 5 operably linked to one or more control sequences or a vector of claim 8 or 9 stably into the genome of a plant cell.

19. A method for the production of transgenic plant cells, plant tissues or plants comprising the introduction of a nucleic acid as defined in any of claims 1 to 5 in an expressible format or a vector of claim 8 or 9 in said plant cell, plant tissue or plant.

20. A method according to claims 18 or 19 further comprising regenerating a plant from said plant cell.

21. A transgenic plant cell comprising a nucleic acid as defined in any of claims 1 to 5 which is operably linked to regulatory elements allowing transcription and/or expression of said nucleic acid in plant cells or a transgenic plant cell obtainable by a method of claim 19.

22. A transgenic plant cell of claim 21 wherein said nucleic acid as defined in any of claims 1 to 5 is stably integrated into the genome of said plant cell.

23. A transgenic plant or plant tissue comprising plant cells of claim 21 or 22 or a transgenic plant obtainable by the method of claim 20.

24. A harvestable part of a plant of claim 23.

25. The harvestable part of a plant of claim 24 which is selected from the group consisting of seeds, leaves, fruits, stem cultures, rhizomes and bulbs.

26. The progeny derived from any of the plants or plant parts of any of claims 23 to 25.

27. A method for identifying compounds or mixtures of compounds which specifically bind to a polypeptide of claim 12 or 13, comprising the steps of
(a) combining a polypeptide of claim 12 or 13 or a cell expressing said polypeptide with said compound or mixtures of compounds under conditions suitable to allow complex formation, and,
(b) detecting complex formation, wherein the presence of a complex identifies a compound or mixture of compounds which specifically binds said polypeptide.

28. The method of claim 27, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:
(a) detection of binding by direct detection of test compound/polypeptide binding;
(b) detection of binding using a competition binding assay; and
(c) detection of binding using an assay for testing the activity of the DIM1- interacting molecule.

29. A method for identifying and obtaining compounds interacting with or modulating the activity of a polypeptide of claim 12 or 13 comprising the steps of:
(a) providing a two-hybrid system wherein a polypeptide of claim 12 or 13 and an interacting protein partner, preferably a DIM1 molecule are expressed,
(b) interacting said compound with the complex formed by the expressed polypeptides as defined in a), and,
(c) performing measurement of interaction of said compound with said polypeptide or the complex formed by the expressed polypeptides as defined in (a).

30. A method for modulating the activity of a polypeptide of claim 12 or 13 comprising contacting a polypeptide of claim 12 or 13 or a host cell of claim 10 or 11 expressing said polypeptide with a compound which binds to the polypeptide or obtainable by any of the methods of claims 27 to 29, in a sufficient concentration to modulate the activity of the polypeptide.

31. A method for modulating the growth of a plant, comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the growth of the plant, thereby modulating the growth of the plant.

32. A method for modulating the cell cycle in a plant, comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the cell cycle in the plant, thereby modulating the cell cycle in the plant.

33. A method for enhancing overall growth and yield of a plant comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the growth of the plant, thereby enhancing overall growth and yield of said plant.

34. A method for modulating pre-mRNA splicing in a plant cell comprising introducing into the cell a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate pre-mRNA splicing in the cell, thereby modulating pre-mRNA splicing in the cell.

35. A method for modulating vesicle transport/processing in a plant cell comprising introducing into the cell a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate vesicle transport/processing in the cell, thereby modulating vesicle transport/processing in the cell.

36. A method for increasing yield of a plant comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the growth of the plant, thereby increasing yield of said plant.

37. A method for enhancing stress tolerance in a plant comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to modulate the growth of the plant, thereby enhancing stress tolerance of said plant.

38. A method for conferring enhanced resistance to pathogens of a plant comprising introducing into the plant a DIM1 interacting (DIMIC) molecule or a DIMIC modulator in an amount sufficient to confer resistance to pathogens of the plant, thereby conferring enhanced resistance to pathogens of said plant.

39. A method according to any of claims 31 to 38 wherein at least one nucleic acid encoding a plant DIM1 interacting (DIMIC) molecule, a homologue or a derivative thereof or an enzymatically active fragment thereof is expressed in specific cells or tissues of said plant.

40. A method according to claim 39 comprising stably integrating into the genome of said plant or in specific plant cells or tissues of said plant at least one expressible nucleic

acid encoding a DIM1 interacting (DIMIC) molecule, a homologue or a derivative thereof or an enzymatically active fragment thereof

41. A method according to claim 39 or 40 wherein said expression of said nucleic acid leads to overexpresion of a DIM1 interacting (DIMIC) molecule in said plant.

42. A method according to claim 39 or 40 wherein said expression of said nucleic acid leads to downregulation of expression of a DIM1 interacting (DIMIC) molecule.

43. A method according to any of claims 31 to 42 wherein said DIM1 interacting (DIMIC) molecule is selected from any of the nucleic acids as defined in any of claims 1 to 4.

44. A method according to any of claims 31 to 42 wherein said DIM1 interacting (DIMIC) molecule is selected from the group of nucleic acids as given in any of SEQ ID NOs 35 to 48 or 94, or a homologue or a derivative thereof, or a functional fragment thereof.

45. A method according to any of claims 31 to 44 wherein the expression or activity of a nucleic acid encoding a plant DIM1 interacting (DIMIC) molecule or a homologue thereof is modulated by a DIMIC modulator.

46. A method according to claim 45 wherein said DIMIC modulator is selected from the group consisting of an antibody of claim 16, an antisense molecule of claim 4, a ribozyme, or a compound obtainable by any of the methods of claims 27 to 29.

47. A method of claim 46 wherein said DIMIC modulator is capable of modulating DIMIC nucleic acid expression.

48. A method of claim 46 wherein said DIMIC modulator is capable of modulating DIMIC polypeptide activity.

49. A method according to any of claims 31 to 48 comprising co-expression of a DIM1 interacting (DIMIC) molecule or a DIMIC modulator and a DIM1 molecule in said plant.

50. A method for detecting the presence of a polypeptide of claim 12 or 13 in a sample comprising:

(a) contacting the sample with a compound which selectively binds to said polypeptide; and

(b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 12 or 13 n the sample.

51. A method of claim 50, wherein the compound which binds to the polypeptide is an antibody.

52. A method for detecting the presence of a nucleic acid molecule as defined in any of claims 1 to 4 in a sample comprising:
    (a) contacting the sample with a nucleic acid probe or primer of claim 6 or 7 which selectively hybridizes to or amplifies the nucleic acid molecule of any of claims 1 to 4, and
    (b) determining whether said nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid molecule of any of claims 1 to 4 in the sample.

53. The method of claim 52, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

54. A diagnostic kit comprising at least one of the nucleic acid molecules of claims 1 to 5, the polypeptides of claims 12 to 14, the antibodies of claim 16, the compounds obtainable by the method of any of claims 27 to 29.

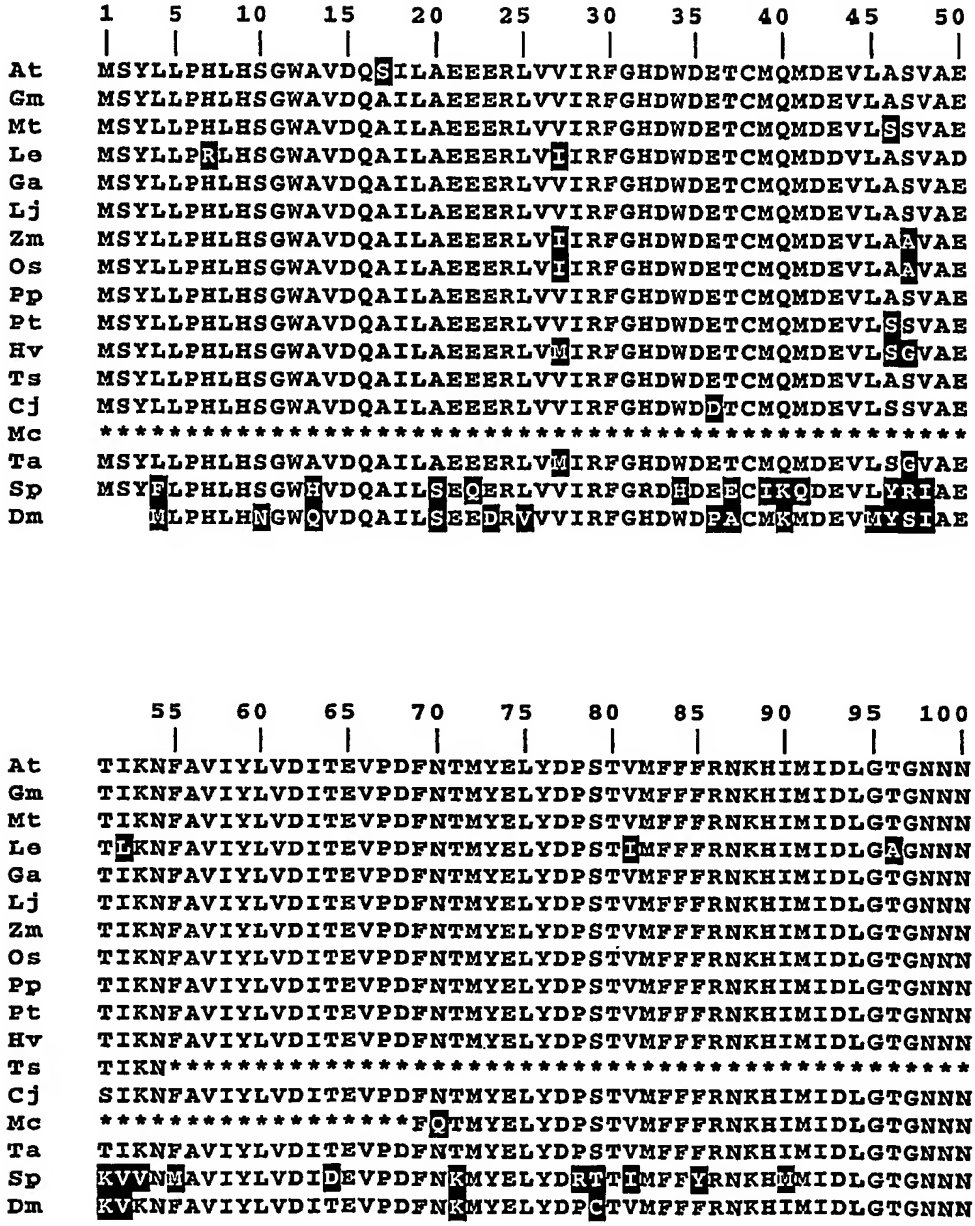
1 5 10 15 20 25 30 35 40 45 50
At MSYLLPHLHSGWAVDQSILAEEERLVVIRFGHDWDETCMQMDEVLASVAE
Gm MSYLLPHLHSGWAVDQAILAEEERLVVIRFGHDWDETCMQMDEVLASVAE
Mt MSYLLPHLHSGWAVDQAILAEEERLVVIRFGHDWDETCMQMDEVLSSVAE
Le MSYLLPRLHSGWAVDQAILAEEERLVIIRFGHDWDETCMQMDDVLASVAD
Ga MSYLLPHLHSGWAVDQAILAEEERLVVIRFGHDWDETCMQMDEVLASVAE
Lj MSYLLPHLHSGWAVDQAILAEEERLVVIRFGHDWDETCMQMDEVLASVAE
Zm MSYLLPHLHSGWAVDQAILAEEERLVIIRFGHDWDETCMQMDEVLAAVAE
Os MSYLLPHLHSGWAVDQAILAEEERLVIIRFGHDWDETCMQMDEVLAAVAE
Pp MSYLLPHLHSGWAVDQAILAEEERLVVIRFGHDWDETCMQMDEVLASVAE
Pt MSYLLPHLHSGWAVDQAILAEEERLVVIRFGHDWDETCMQMDEVLSSVAE
Hv MSYLLPHLHSGWAVDQAILAEEERLVMIRFGHDWDETCMQMDEVLSGVAE
Ts MSYLLPHLHSGWAVDQAILAEEERLVVIRFGHDWDETCMQMDEVLASVAE
Cj MSYLLPHLHSGWAVDQAILAEEERLVVIRFGHDWDDTCMQMDEVLSSVAE
Mc **********
Ta MSYLLPHLHSGWAVDQAILAEEERLVMIRFGHDWDETCMQMDEVLSGVAE
Sp MSYFLPHLHSGWHVDQAILSEQERLVVIRFGRDHDEECIKQDEVLYRIAE
Dm MLPHLHNGWQVDQAILSEEDRVVVIRFGHDWDPACMKMDEVMYSIAE
55 60 65 70 75 80 85 90 95 100
At TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Gm TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Mt TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Le TLKNFAVIYLVDITEVPDFNTMYELYDPSTLMFFFRNKHIMIDLGAGNNN
Ga TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Lj TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Zm TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Os TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Pp TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Pt TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Hv TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Ts TIKN**********
Cj SIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Mc **********FQTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Ta TIKNFAVIYLVDITEVPDFNTMYELYDPSTVMFFFRNKHIMIDLGTGNNN
Sp KVVNMAVIYLVDIDEVPDFNKMYELYDRTTIMFFYRNKHMMIDLGTGNNN
Dm KVKNFAVIYLVDITEVPDFNKMYELYDPCTVMFFFRNKHIMIDLGTGNNN

FIGURE 1

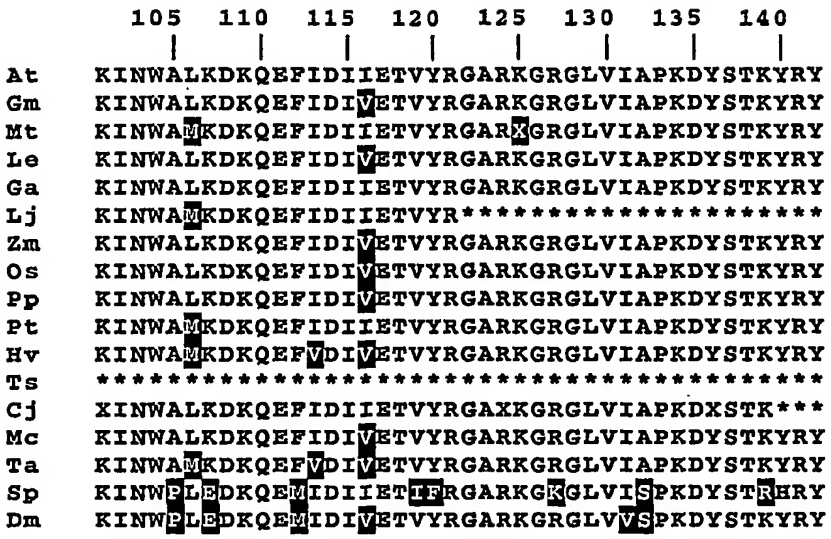

FIGURE 1 (continued)

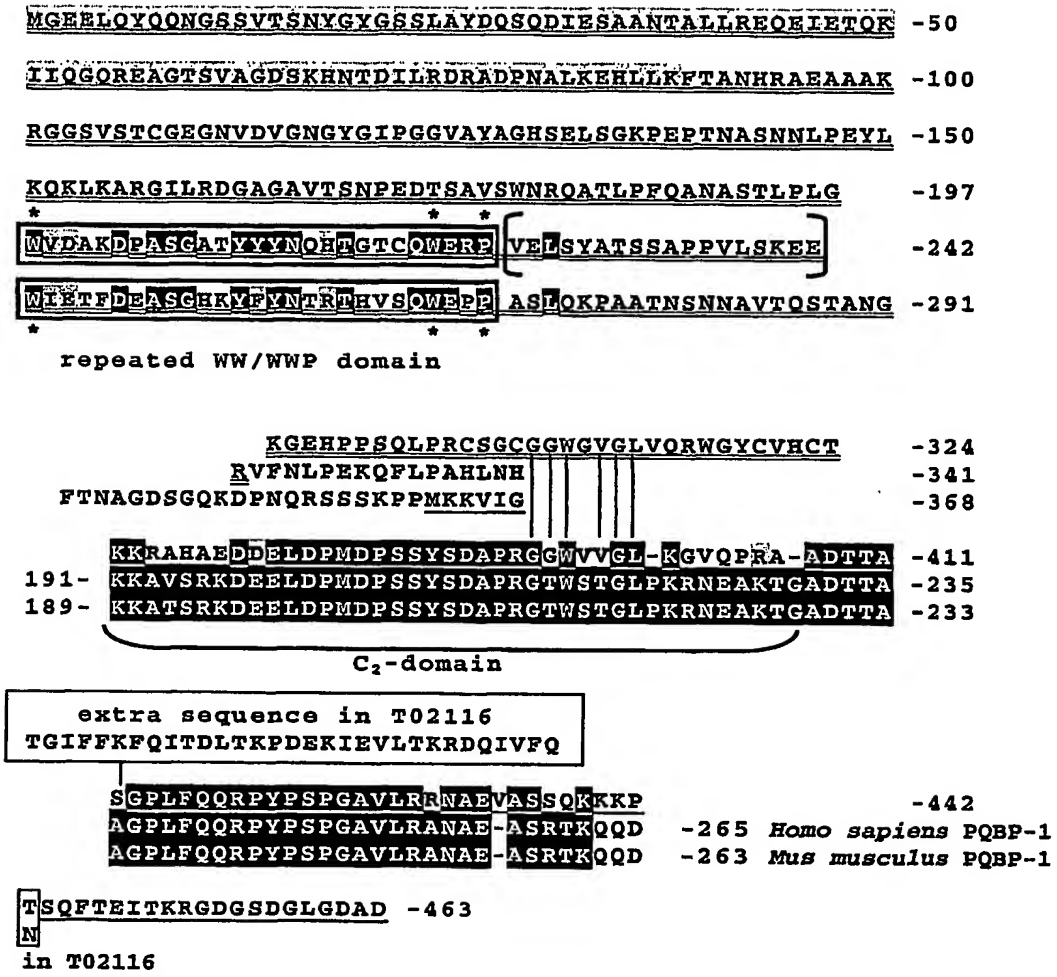
MGEELQYQQNGSSVTSNYGYGSSLAYDQSQDIESAANTALLREQEIETQK -50
IIQGQREAGTSVAGDSKHNTDILRDRADPNALKEHLLKFTANHRAEAAAK -100
RGGSVSTCGEGNVDVGNGYGIPGGVAYAGHSELSGKPEPTNASNNLPEYL -150
KQKLKARGILRDGAGAVTSNPEDTSAVSWNRQATLPFQANASTLPLG -197
WVDAKDPASGATYYYNQHTGTCQWERP VELSYATSSAPPVLSKEE -242
WIETFDEASGHKYFYNTRTHVSQWEPP ASLQKPAATNSNNAVTQSTANG -291
repeated WW/WWP domain
KGEHPPSQLPRCSGCGGWGVGLVQRWGYCVHCT -324
RVFNLPEKQFLPAHLNH -341
FTNAGDSGQKDPNQRSSSKPPMKKVIG -368
KKRAHAEDDELDPMDPSSYSDAPRGGWVVGL-KGVQPRA-ADTTA -411
191- KKAVSRKDEELDPMDPSSYSDAPRGTWSTGLPKRNEAKTGADTTA -235
189- KKATSRKDEELDPMDPSSYSDAPRGTWSTGLPKRNEAKTGADTTA -233
C2-domain
extra sequence in T02116
TGIFFKFQITDLTKPDEKIEVLTKRDQIVFQ
SGPLFQQRPYPSPGAVLRRNAEVASSQKKKP -442
AGPLFQQRPYPSPGAVLRANAE-ASRTKQQD -265 Homo sapiens PQBP-1
AGPLFQQRPYPSPGAVLRANAE-ASRTKQQD -263 Mus musculus PQBP-1
TSQFTEITKRGDGSDGLGDAD -463
N
in T02116

FIGURE 2

4/21

```
17241                                 atgggagaag agctgcaata tcagcagaat ggttcatctg
17281 tgacatctaa ttatggttat ggaagttcgc ttgcttacga tcaatctcaa gacattgaat
17341 ctgccgcaaa cactgctctt ctccgagaac aagtacactt tttttgtct tcttcttctt
17401 caatataagc ttctaaattc cattcccgga ctttgccaat ggttaccatg ttgatttggt
17461 tttgtgtgtt ttcaggaaat agaaacgcag aagattatac aaggtcaaag gttctgtctt
17521 tctcttggtt tgttctttat atccatctgt tatatatgct taatgcttgt attgatcatc
17581 ttgcttgtttc tcattcttgg gcaatgagaa atttcgtgtg tttaggtgct ttgtctaaca
17641 atttgtgctt aattggaaac gttctcctat tttcagagaa gctgggactt cagtagccgg
17701 agactctaag cataacactg atattcttcg agatcgtgct gaccccaatg ctctaaaggt
17761 tctgattcac tattttttc gttatcatc aagcggctat tggcttaata agtcctataa
17821 tcacatgaag ttttgataca tattatcttt atgcaggaac atttgctaaa gttcactgca
17881 aatcatagag cagaagcagc ggcaaagcgt ggtgggagtg tgagtacatg cggagaaggt
17941 tagtcataat agcttacgag ctcagtcatt taggaaaatt caggaaactg atccaattat
18001 tcgatttatt ccgttctagg taatgtggat gttggaaacg gttatggtat accgggtgga
18061 gttgcatacg cgggtcactc cgagctcagt gggaagcctg aacctaccaa tgcttcaaac
18121 aatttgcctg aatatctcaa acaaaaatta aaggccaggg gaattctcag agatggtgca
18181 ggggctgtta catccaaccc agaagatgta attattcatc cctcagactg taatccaggg
18241 tgacatatgt ctgcaatgac ttgtccatat ttcagccct ttgtttcagt ttatttatt
18301 aatatttgtt aacctttga cgttttgtt tgcttgtgtt tcttctcttt tccagacatc
18361 agcggttagt tggaatagac aggcaacttt accttttcag gctaatgcaa gcaccttacc
18421 attaggatgg gtatgaatga gatacactta aaagtctttt cttcttcatt ttgtcttg
18481 gtttactatc tggtgcagca atattttctg ctgtgaaagt gcggtaatct tttgagattc
18541 tctaaatgca tgtgataatg tggtcataca caaatagtct ttagatctca atccttaagg
18601 atgaccactt atggttctga tgatgtgttt atagttgtta taccacttaa atttgttctg
18661 tttcattcag gtagatgcaa aagatcctgc cagtggtgct acctattact ataaccagca
18721 cactggaaca tgccagtggg aaaggcctgt tgagctttct tatgccactt caagtgcacc
18781 acctgtactt tctaaagaag agtggattga aacatttgat gaagcatctg gtatgatttt
18841 cacacattta gtcatcataa tgactagctc gttgctacaa tattcatatc aaagattata
18901 aaagtatgat atctttctcc tttcttatt gtggtattat gctgctagca aaatgaattt
18961 aatagcttat gatctttttc atcttggaaa gaaacttaca tcatcacatg ttagaaataa
19021 gaatgaggaa atccttgtgt tttctgctgg aagttttgc ctcttattga ccagcttctt
19081 ctgatttctg ttgtactgta cacgcaatgc ttgcttaatc tttggatctg ttggttgcat
19141 ctttgaactc tgaacctttt gattcctatg cacaggccat aagtatttct acaatacaag
19201 gacacatgtg tctcagtggg aacctccagc ttctttacag aagcccgctg ccacaaactc
19261 taacaacgct gttacccaaa gtacggctaa tgggaagggg gagcatcctc catctcagct
19321 gccaagatgc agcggatgtg ggggctgggg agtgggcctt gtccagagat ggggttattg
19381 tgttcattgt accaggtagg aacccttcca tttcttgtta tgcttttga gttagattgg
19441 tgcaaagaga ctgattcggt tgaagactgt taataatgta tatacctcta gtcctcaacc
19501 atgactccct tttattact ggctcagatt tgtgttatag gaagtattta ttagtgtgca
19561 tcaagaccct aaatacactt tgtttacttg catattttga ggcttaagcc cctctgcctt
19621 tggttaacgg gacttatatt atatgggatt cagggttttc aatctcccag aaaagcaatt
19681 cttgccagca catttaaacc atttcacaaa tgcaggcgat tctggccaaa aggatccgaa
19741 tcagaggtac gcatctctca aaatcttaga tttagaaaaa atattcttct atgccgtca
19801 tcttttctaaa atacatagtg ttccataaga gagctattgg tggactctgg tcatcttctt
19861 aagatctgtt accagatatt tagacaaccc atttggatac aaatgttggc cctcttttcg
19921 gagggacata atttctgtta aaaagcattc cattatcag cagtaaaatg aagatggaac
19981 acattacatg tagtagtgaa ccaggaggtt cccattgact atgtttacat tgtacttact
20041 cttgtggtat attctttaca catgatggaa aaatgaaaat ggttgaaaca ggtccagttc
20101 aaaacctccc atgaaaaaag taataggcaa aaagcgtgct catgccgaag atgatgagtt
20161 ggatccaatg gacccgagct cttactcaga tgctccacgg ggaggctggt aatgtccata
20221 acacccattt taaccacttc atgaaacggc atgacctata atactgttct ctatctctct
20281 gtgctatgtc atctgaattg aggagaacca agttctttt tggttaacaa ccaagctaat
20341 tatgctcata ggaatcataa tatcaggttg atctgattgg ttctcatctt gtagggttgt
20401 tggactgaaa ggagtacaac cgcgagccgc tgatacaact gcttcggtaa tgggatttac
20461 cttttcagac tgggatcttc tttaaattcc aaataaccga ctcactaaa ccagatgaaa
20521 aaattgaggt tctaactaaa cgagatcaaa tcgtgtttca gggtcctctg tttcaacaac
20581 gaccatatcc atcacctgga gctgttctga ggagaaacgc agaggtggca tcatcacaga
20641 agaagaaacc aaactctcag ttcacagaaa ttacaaagag aggcgatgga agtgatggtc
20701 ttggagacgc agattga
20718 gtg agagcttccc tcatcttata agttggttac cgtgtcttct
20761 tgtgctttgt ctgttgcttt ctccatatat gagatcaagg agagtagttt gtttactttt
20821 tagtgaatac aaacaaacca tacttcttca tgtgaatttt ttttaacacc agattcaaag
20881 tatgtataca tataattact cagtagcaag caacggtctt caac [aaaaaa aaaaaaa]
```

FIGURE 3

```
100439                                                              at
100441 gggaatacct gatggttcac tgctagatct aattgataag gttaggtctt ggataacttc
100501 ggattcaagt gattctttgt tcttgttgtc ttcatctaaa caagactttg gaattatgcc
100561 tatcgtttcc aaaatgtgtc atgactgtgg aacaaaggtg gaacaaggat attgttgtct
100621 tagttgtggt tcttgttggt gcaagagttg ttctgataca gaagagagca agatgaagtt
100681 gtgtagagaa tgtgatgctg aggttcgtga gttaagggtt aagagctatg ataaggttca
100741 tcctcgggat agccctgatc cgccttcttc tcttgctacc gaaagcgaga gtcttgcctc
100801 gagtcttgaa atccgtgatt gtaggaacat ggcttctatt cgttgctatc ctagcaggta
100861 aagtttggat ttttacattc aattatgtgc gaaaaagatt caatttttat gttcaattat
100921 gagtctttaa acatgttgga gtaagtgaat gatttgtgtt tgcagggag aggaagaaga
100981 ggctagatat tgtggtaaac agttgttaag cccttcgagt gataactatc aggacagttc
101041 agatatagag tcgggtagtg ttagtgctag acatgagctt tttagttgta aatcttctgc
101101 tgggtccagt cctcacgaca gtccattaag gaacaatttt agtcctcttg ggcgctttgt
101161 acagcacgcg aaagacctaa gaagccctac tgtttgttct tttgataatc atcaagaaca
101221 actcttggct gataatttag ttaagccagg tcagggggta ttggaacagg aagatcacga
101281 ggaggaagag gataagttgc aacagccgtt ggattttgag aacaacggcc gcatctggta
101341 tcctccacca ccggaagatg agaatgatga tgctgagagt aattactttc attatgatga
101401 cgaagatgac gatattgggg actctgctac cgagttttca ttgagtagta gcttttctag
101461 ccacattcct accaaagaga aattgggaga aaacagcaat gaacctcttc gaaccgtggt
101521 gcatgaccac ttccgagctc ttgttgcaga gctattacgt ggggaggagc ttagtccatc
101581 tgatgatggc agtgccgggg agtggctgga tattgtcact gctttagcgt ggcaagctgc
101641 aaatttgta aagccggata ctcgtgcagg cggcagtatg gatcccggga attatgtaaa
101701 aataaaatgt gtagcatctg gaaatcaaaa tgagaggtat aactagttgt tgaaacctg
101761 tttaaatcgt cctctgactg atgattacta atttgggtga tagttactaa atgaatacct
101821 tgtttctgcg gcattctcat caggggaata gtgtgtagca agaacattac gcacaagcga
101881 atgatatcgc agtacaaaaa tccaagggtg atgcttttag ctggttctct tgagtatcag
101941 agagttgcag gccaattagc atcctttaac acgttgcttc aacaggttgg tgtgttactg
102001 tttctttatg atggctgcat aaaatctctg attaacaaac tgaacaactg aattatgtgg
102061 gttggacagg aaaatgaaca tatgaaggcg ataatagcaa aaatagagtc ccttcgtcct
102121 aatgttctgc tggtagagaa aagtgcatct tcatatgccc agcaatacct tctagagaag
102181 gaaatttcat tggtgctcaa tgtgaaaagg tcattgctgg accgtatagc ccgttgcact
102241 ggtgctgttc tttgcccctc actagatagt attagtactg ctcgattggg tcactgtgag
102301 ctgttccgga ctgagagagt gttggagcaa catgaagctg gcaatcagtc aaacagaaaa
102361 ccatcaagga cattgatgta ctttgaaggg tgtcctaggc gcctaggttg cacggtatgc
102421 tcctttattg gtttcacagt tggtgttttg catagcttga acctttgtag ctaagaagtg
102481 tgctatctta aatttattaa cttaccttac tgactttctg aaacttgtgc aggttgtgct
102541 tcggggaagt tgtcgtgagg agctaaagaa ggttaaacat gttattcagt acgctgtatt
102601 tgcagcttat cacttatcgt tggagacttc attccttgcg gatgaaggtg cttctctgcc
102661 taaaattaga ctgaagcaac cagggatggt gcggactgca tcacaaagga ggataattga
102721 tgaaggcatt tccttaataa ctcaatctcc tacagaaaca gatagtcaag ctttgcttga
102781 aacagctgca catgaagatg aacatacggc tcctatgcct gagcatgaag tttgtgaatc
102841 attgtgtgag gattttgatc ctactcagat attcccaccg tcttctgaag ttgagactga
102901 acaatctgat accctgaatg gagattttgc caataatttg gtaacaagat catattcttc
102961 gaatcagtta aatgatttgc atgaacccac tttgtgctta agtagtgaaa ttcctgaaac
103021 gccgactcaa cagccgagtg gtgaagaaga taacggcaga ggtgaagaag agaaccaatt
103081 agtgaacccc caagatttgc cacaacatga gagcttctat gaagatgatg tctctagtga
103141 gtatttctct gctgcggact ctcatcagag catattggtc tcattttcaa gccgttgtgt
103201 tttgaaagaa tcagtatgtg aacgttcacg actcttgcgg attaagttct atggatcctt
103261 tgataaacct ctcggaagat atttgaaaga tgatcttttc gataaggtag ctctttgatc
103321 tataaagtta gttaatccac ttttcatgtc ttcttaaatt ttgaagtcta atgccttctc
103381 gttccatttc tggacagact tctagttgta gatcctgtaa ggagctggtt gatgcgcatg
103441 tactctgcta ttctcatcag aatggaaacc ttactatcaa tgttagacgt ctcccttcca
103501 tgaagcttcc tggtgaacaa gatgggaaga tatggatgtg gcatcgttgc ttaagatgtg
103561 cacatgtaga tggagtccca cctgctactc gtagagtagt tatgtctgat gcagcatggg
103621 gattgtcttt tggaaagttt ctggaactta gcttttcgaa tcatgcaact gcgaatcgtg
103681 ttgcatcttg cggtcattcg cttcaaagag actgccttcg attctatggg taaacgcagg
103741 ctgtataata tcattctctt acagttatct caatgctgtt acactaataa tattatgatg
103801 tgttttgtag atttggtaac atggtggcat tctttagata ttctcccatc aatatactta
103861 ctgtactcct cccaccgtcg atgcttgagt tcaatagcca tcctcaacag gagtggatac
103921 gaacagaggc agctgaggta gattttaaga aaaacttaag gctcttttgc ttcaggtttt
103981 ggatatgtaa atgctaaaac tcaaatactt ctctgtttgc atttacagc ttgttggtaa
```

FIGURE 4

104041 aatgaggact atgtatacgg agatatctga tatgctcaac cgcatggaag aaaaaagcag
104101 tttgcttgaa cctgaacaat ctgaagcatg tgatctacat agtcgcatca tcggattaat
104161 agatcaactt gtgaaagaaa aagatgaata cgatgtaagc ttctttctgtt tctgaaatct
104221 tgatgatgct tgttgaatat gtatttgatg gaatattcct catcgtgaac aggatgcact
104281 gcaacctatt tttgaggaga atctgcaaat tcagggaagt ttggatatcc tagagctgaa
104341 tcgattgaga cgagcactca tgattggtgc tcatgcttgg gatcatcagc tttacttgtt
104401 aaactctcaa ctcaagaaag catcagtctt taaaactggt gatgacaatg ctcctcggaa
104461 tccagagatg catgatcctc caaagattga ccgtagaatg caggaaggtt cagatgagag
104521 ggatgaacaa tcccatactg attcagaagc taatggtgat aacaaagatc cggaaaacat
104581 tccctcccct ggcacttctc tatctgagag gatagattcg gcgtggttgg gctcgtttca
104641 aaatctagag aaagccgaga ctattgctga gacggaaggc ttttcagctg ttaattcttc
104701 tctccggagg cttgccaggc caattcgagt ccagtctttt gattcagcga tacggtttca
104761 agaacgtatc caaaaaggtt tgccaccatc ttctttgtat ctctcgaccc tcagatcttt
104821 ccacgcttct ggcgagtata ggaatatggt gagggaccct gtttcaaatg tgatgaggac
104881 ttactcacag atgttgccac tggaagtaca gaaattggat ctgattgttg gttcagcgcc
104941 aacatacatc tcttctgcat ctcaaatggc tgatggggca cggatgttga ttcctcaacg
105001 gggccttaac gatatcgtgg ttcctgtata cgatgatgat ccagctagtg ttgtatcata
105061 tgctatcaat tcaaaggaat ataaggagtg gatcgttaac aaaggtcttg cttctagtag
105121 tagtagcagc aatttgaata acagagagtc tgaaccttcc gccttctcga cttggcgttc
105181 tctgtcgatg gatgttgatt acattcaaca cgcggtgtat ggatcttctc aagacgatag
105241 aaagtctccg catttgacca tttctttcag cgaccgcgct tcttcttcct ctactgccac
105301 tgaaggtaaa gtgaagttct ctgtgacgtg ttattttgca acgcagtttg acactctgag
105361 aaagacttgc tgtccgagtg aagtggactt tgtgagatcc ttgagccgtt gtcagagatg
105421 gtctgcacag ggagggaaaa gcaatgttta ctttgccaag tcattggacg agaggttcat
105481 cataaaacaa gtcgtcaaaa ccgagctgga ttctttcgag gattttgcac ctgagtactt
105541 caaatatttg aaggaatcac tcagttccgg gagccctact tgtcttgcta agattctcgg
105601 tatctaccag gttagtgcca atatagggag gaatgtgaaa ttgtagggac ctgaaatcaa
105661 ataagtgagc atgtgaatga ttgatttgga atgatattta tttgaggtct caattaaaca
105721 tccgaaaggt ggaaaagaga caaagatgga tttgatggtg atggagaatc tcttctacaa
105781 cagaagaata tctaggatct atgatctcaa gggatctgca cggtcaaggt acaatccaaa
105841 cacatccgga gcagacaaag ttttgttaga catgaatctg ctagagacac tgcgcacaga
105901 accgatattc ctaggaagca aggccaagag aagcttggaa agagcgatat ggaacgacac
105961 caatttcttg gctgtaagtg ttttcattcag ccccaattca tatttgtttc ttaagtgttg
106021 tgattaatct gttcccgat tatcctcttg tggttagtct gtggatgtaa tggactactc
106081 attgctggtt ggattcgacg aagagcggaa agaactagtt ctcgggatca tcgacttcat
106141 gagacagtac acatgggaca agcacctaga gacttgggtt aaagcttcag gcattttggg
106201 tggacccaag aatgcttctc caactatagt ctcaccaaaa cagtacaaga ggaggtttag
106261 aaaggccatg accacttatt ttctcactgt tcctgagcca tggacctctt ga
106313 gtgaaaga
106321 aaagaaacca tttttcatca ttcttttgtt cactgattcc aatttgttct tctcaaataa
106381 aaagggaaat tggttaacag acacatataa tataatacag gttcgaacct gactgattga
106441 tttatctttt ttttttttt ggtgtgttgg tttctctgag ttctggagaa ttgtatattg
106501 taattatcgg ccaaactaaa gccccccaaa tttaaagaaa aagaatacta atgatttgcc
106561 tttt [aaaaaaaaaaaaaaaaaaaaa]

FIGURE 4 (continued)

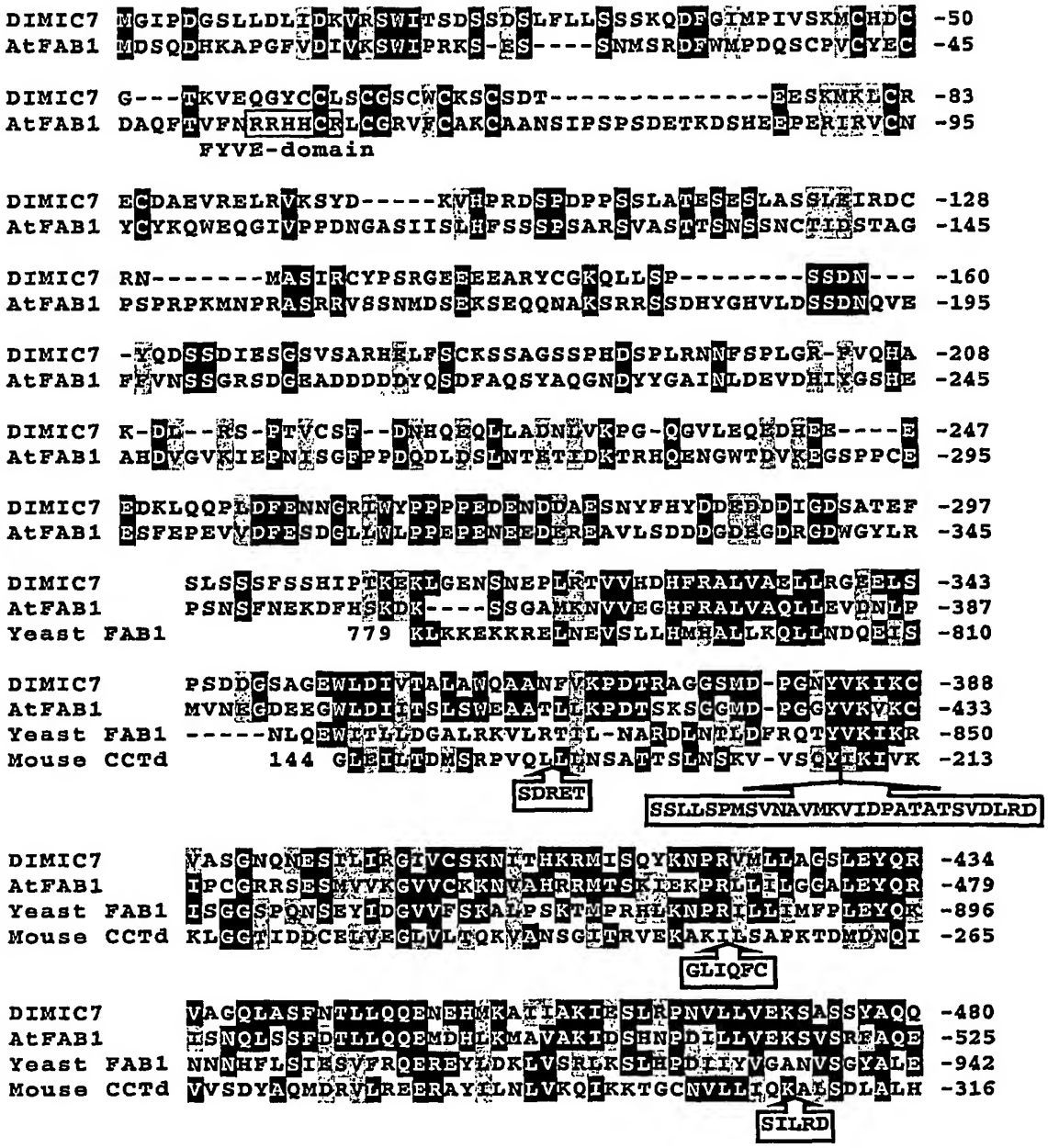
DIMIC7 MGIPDGSLLDLIDKVRSWITSDSSDSLFLLSSSKQDFGIMPIVSKMCHDC -50
AtFAB1 MDSQDHKAPGFVDIVKSWIPRKS-ES----SNMSRDFWMPDQSCPVCYEC -45
DIMIC7 G---TKVEQGYCCLSCGSCWCKSCSDT----------EESKMKLCR -83
AtFAB1 DAQFTVFNRRHHCRLCGRVFCAKCAANSIPSPSDETKDSHEEPERIRVCN -95
FYVE-domain
DIMIC7 ECDAEVRELRVKSYD-----KVHPRDSPDPPSSLATESESLASSLEIRDC -128
AtFAB1 YCYKQWEQGIVPPDNGASIISLHFSSSPSARSVASTTSNSSNCTIDSTAG -145
DIMIC7 RN-------MASIRCYPSRGEEEEARYCGKQLLSP--------SSDN--- -160
AtFAB1 PSPRPKMNPRASRRVSSNMDSEKSEQQNAKSRRSSDHYGHVLDSSDNQVE -195
DIMIC7 -YQDSSDIESGSVSARHELFSCKSSAGSSPHDSPLRNNFSPLGR-FVQHA -208
AtFAB1 FEVNSSGRSDGEADDDDDYQSDFAQSYAQGNDYYGAINLDEVDHIYGSHE -245
DIMIC7 K-DL--RS-PTVCSF--DNHQEQLLADNLVKPG-QGVLEQEDHEE----E -247
AtFAB1 AHDVGVKIEPNISGFPPDQDLDSLNTETIDKTRHQENGWTDVKEGSPPCE -295
DIMIC7 EDKLQQPLDFENNGRIWYPPPPEDENDDAESNYFHYDDEDDDIGDSATEF -297
AtFAB1 ESFEPEVVDFESDGLLWLPPEPENEEDEREAVLSDDDGDEGDRGDWGYLR -345
DIMIC7 SLSSSFSSHIPTKEKLGENSNEPLRTVVHDHFRALVAELLRGEELS -343
AtFAB1 PSNSFNEKDFHSKDK----SSGAMKNVVEGHFRALVAQLLEVDNLP -387
Yeast FAB1 779 KLKKEKKRELNEVSLLHMHALLKQLLNDQETS -810
DIMIC7 PSDDGSAGEWLDIVTALAMQAANFVKPDTRAGGSMD-PGNYVKIKC -388
AtFAB1 MVNEGDEEGWLDIITSLSWEAATLLKPDTSKSGGMD-PGGYVKVKC -433
Yeast FAB1 -----NLQEWITLLDGALRKVLRTTL-NARDLNTLDFRQTYVKIKR -850
Mouse CCTd 144 GLEILTDMSRPVQLLMNSATTSLNSKV-VSQYIKIVK -213
SDRET
SSLLSPMSVNAVMKVIDPATATSVDLRD
DIMIC7 VASGNQNESILIRGLVCSKNITHKRMISQYKNPRVMLLAGSLEYQR -434
AtFAB1 IPCGRRSESMVVKGVVCKKNVAHRRMTSKIEKPRLLILGGALEYQR -479
Yeast FAB1 ISGGSPQNSEYIDGVVFSKALPSKTMPRHLKNPRILLIMFPLEYQK -896
Mouse CCTd KLGGTIDDCELVEGLVLTQKVANSGITRVEKAKILSAPKTDMDNQI -265
GLIQFC
DIMIC7 VAGQLASFNTLLQQENEHMKATIAKIESLRPNVLLVEKSASSYAQQ -480
AtFAB1 HSNQLSSFDTLLQQEMDHLKMAVAKIDSHNPDILLVEKSVSRFAQE -525
Yeast FAB1 NNNHFLSIESVFRQEREYLDKLVSRLKSLHPDIIYVGANVSGYALE -942
Mouse CCTd VVSDYAQMDRVLREERAYILNLVKQIKKTGCNVLLIQKALSDLALH -316
SILRD

FIGURE 5

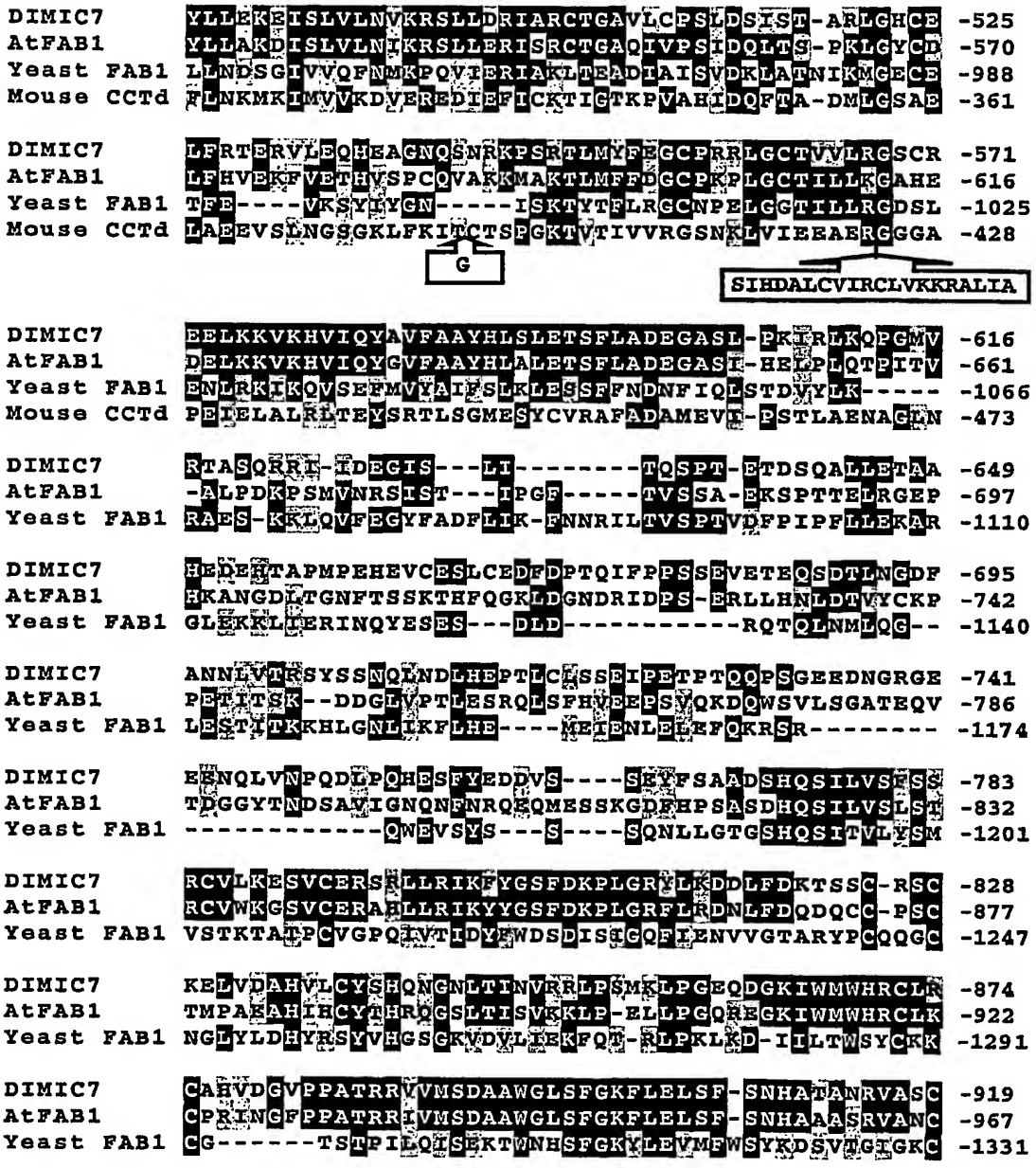

FIGURE 5 (continued)

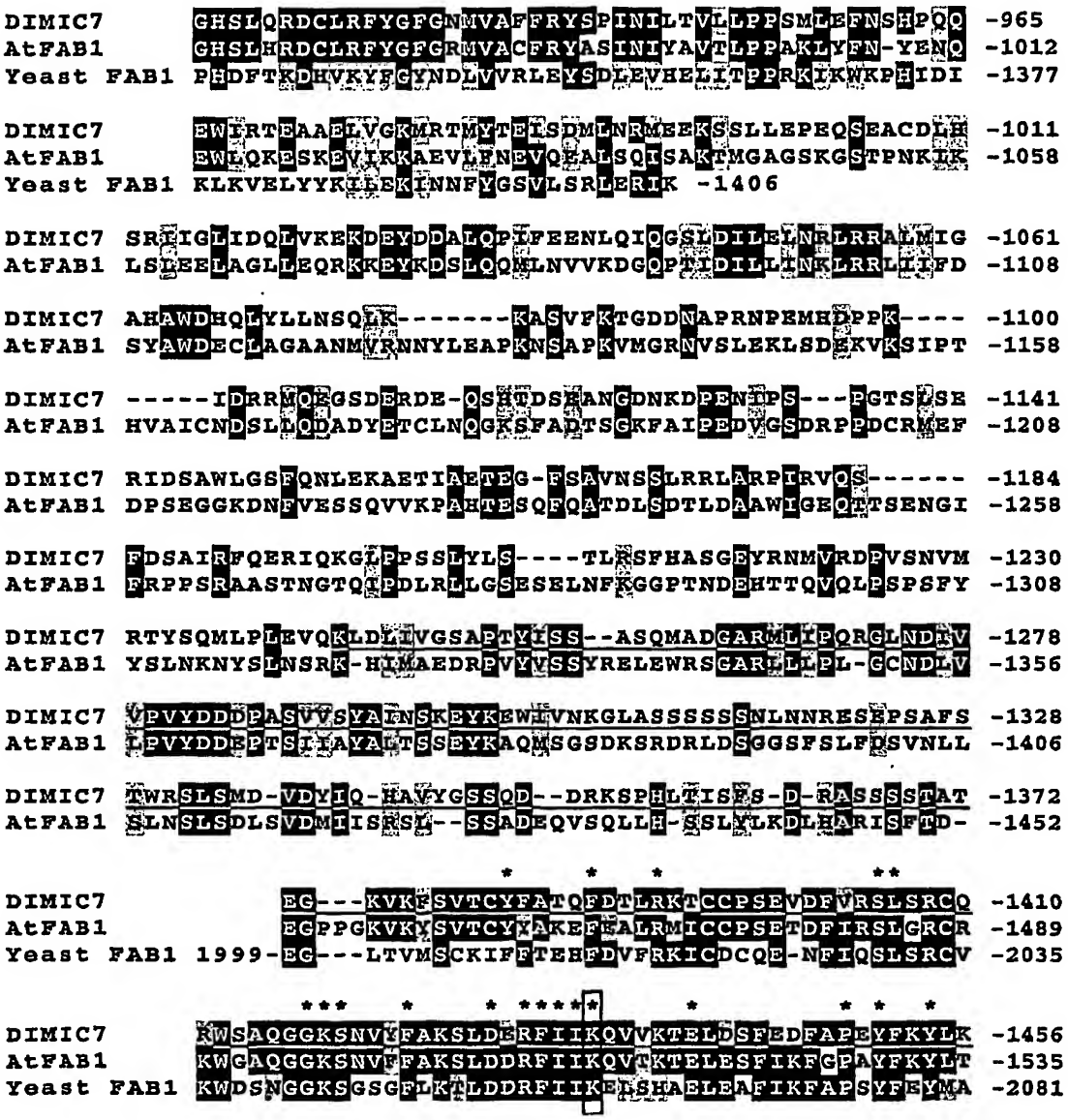

FIGURE 5 (continued)

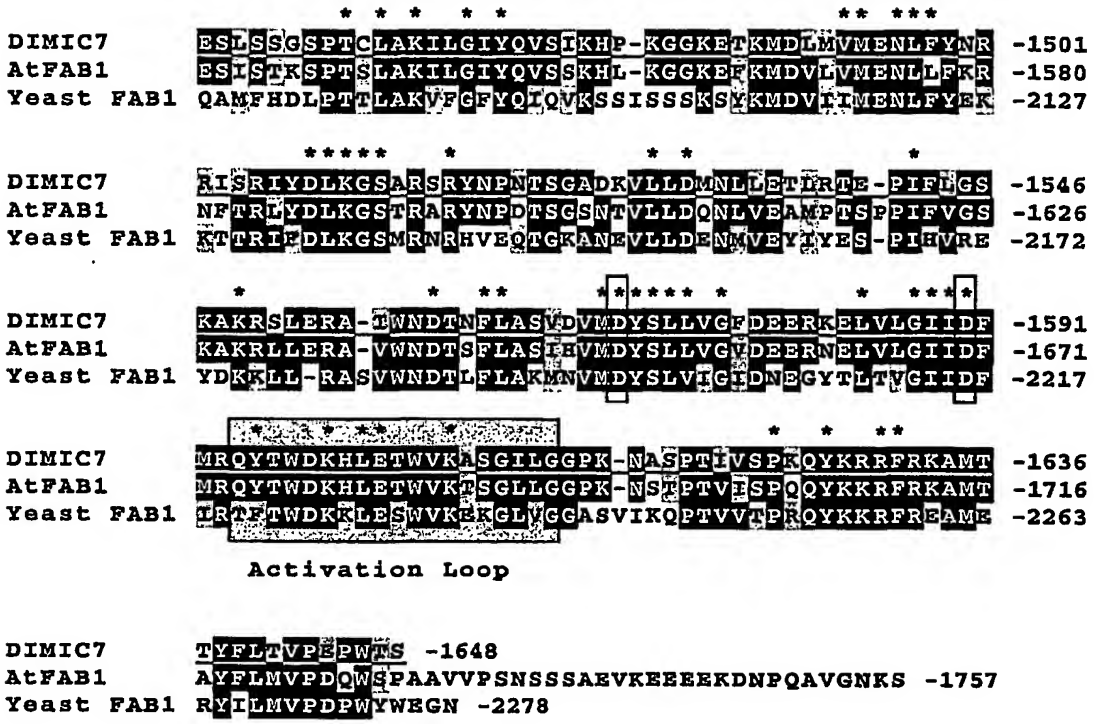

FIGURE 5 (continued)

```
Motif 1 DIMIC7/1
      200 -PLGRFV- 205
      809 -PLGRYL- 814
consensus-PLGR[F/W/Y][M/I/L/V]

Motif 2 DIMIC7/2
      258 -ENNGRIW- 264
      861 -EQDGKIW- 867
consensus-EXXG[R/K/H]IW

Motif 3 DIMIC7/3
      210 -DLRSPTV- 216
      711 -DLREPTL- 717
consensus-DLXXPT[M/I/L/V]

Motif 4 DIMIC7/4
      274 -DDAESNYF- 281
      761 -DDVSSEYF- 768
consensus-DDXXSXYF

Motif 5 DIMIC 7/5
      984 -TEISDMLN- 991
      685 -TEQSDTLN- 692
consensus-TEXSDXLN
```

FIGURE 6

```
21435                                                    atggg aacatcatct
21420 ggttcgaatc atcctcacca aatgttacca ccacgtcaac aacaacgaag tggaggagga
21360 cttgaaacag ctctttcact agtctcttca gatcaagagc cgcgtcgcga gtctccagcg
21300 gagagtgcga gttctcaaga aacatggcca ctaggagaca ctgttgctgg gaagaagtca
21240 atgagtcaaa agacagagcc tgattctatg gaacaaactg tgaatgtgat gcaccatgtt
21180 tcgaatgcgg acaaggtatc ggtacgggat atcgctaggg aaagagttga gttagttgca
21120 gagagaatgc atcgattacc ggatgagttt cttgacgagt taaagaacgg tcttaagtct
21060 atccttgaag ggaatgtagc gcagagtgtt gatgagttca tgttcttgca gaaggttgtt
21000 cagagtagaa ctgatttgag ctctgtaaca cttgttagag ctcatcgggt gcagcttgag
20940 atacttgtag cgataaatac cgggattcag gcgtttttgc atccgaatat tagtctgtct
20880 cagccgtctc ttatcgagat ttttgtatac aagagatgta gaaacatagc ttgccaaaat
20820 caactacctg ctgatgattg ctactgtgat atatgtacta ataggaaagg gttttgtaac
20760 ctttgtatgt gtacgatctg taacaagttt gatttttcgg ttaatacttg ccgctggatc
20700 ggatgtgact tgtgttccca ctggactcat acggattgtg ctattcggga tggacagatc
20640 accactggat cctctgctaa gaataatact tcaggtccag gagagatagt gtttaagtgt
20580 cgggcttgta accgaacgtc tgagctgttg ggatgggtta aagatgtgtt tcagcactgt
20520 gcaccgaact gggatcgaga atctttgatg aaggaacttg attttgttag taggattttc
20460 cgcggaagcg aagatcaacg aggtaggaaa ctgttttgga agtgtgagga gcttattgac
20400 aagattaaag gtggtttggc tgaagcaact gctgccaaat tgatattaat gtttttccaa
20340 ggtaacttct ttttatgaaa ccaacactac tatttgtctg cagagtcttg aacgttttcg
20280 gttactgttt tacctaagag atcgaatcag actctgcgaa gagctttgaa aatggagaag
20220 gtggacgttt gatggcacct caggatgcat gcaaccgaat tgctgaagtt gtacaagaga
20160 ctttgaggaa aatggagata gtggctgagg agaagatgag aatgttcaag aaggcacgca
20100 tggctcttga gacatgcgat agagagctcg aagacaaagc caaggaagtt tcagaactga
20040 aagcagagag gcagaagaag aaacttcaga tagatgagct agagagaatc gtgaggctga
19980 agcaagcgga ggcagacatg ttccagctta aagcaaatga agcaaagcgg gaggctgata
19920 ggttgcagag gattgtacta gcgaaaatgg ataagtctga ggaggaatac gcaagtaact
19860 atctgaaaca gaggctgagt gaggccgagg cagagaagca gtatctgttt gagaagatta
19800 agctgcagga aaactcaagg gtggcatcac aaagcagtgg tgggggagga gacccgtcac
19740 aggtgatgat gtactcaaag atccgcgatc tgcttcaagg atacaatctc tctcccaagg
19680 ttgatcctca attaaacgag cgtaatcctt tcagatccaa tccctaa
19633 tgc ttcttcctcc
19620 cttatatata tatgtagttt tcatatccta aatttcatca ctgtgttttg taaaacgagt
19560 gtttgttgcc ggtaggatct atctatatta ctgatgtttt gtaatttgtg tgttctcact
19500 ttgttgatat gacgttgttt atctatgtta tggtttcttg gttggttg [aaaaaaaaaaa
      aaaaaaa]
```

FIGURE 7

A

```
MGTSSGSNHPHQMLPPRQQQRSGGGLETALSLVSSDQEPRRESPAESASS -50
QETWPLGDTVAGKKSMSQKTEPDSMEQTVNVMHHVSNADKVSVRDIARER -100
VELVAERMHRLPDEFLDELKNGLKSILEGNVAQSVDEFMFLQKVVQSRTD -150
LSSVTLVRAHRVQLEILVAINTGIQAFLHPNISLSQPSLIEIFVYKRCRN -200
IACQNQLPADDCYCDICTNRKGFCNLCMCTICNKFDFSVNTCRWIGCDLC -250
SHWTHTDCAIRDGQITTGSSAKNNTSGPGEIVFKCRACNRTSELLGWVKD -300
VFQHCAPNWDRESLMKELDFVSRIFRGSEDQRGRKLFWKCEELIDKIKGG -350
LAEATAAKLILMFFQEIESDSAKSFENGEGGRLMAPQDACNRIAEVVQET -400
LRKMEIVAEEKMRMFKKARMALETCDRELEDKAKEVSELKAERQKKKLQI -450
DELERIVRLKQAEADMFQLKANEAKREADRLQRIVLAKMDKSEEEYASNY -500
LKQRLSEAEAEKQYLFEKIKLQENSRVASQSSGGGGDPSQVMMYSKIRDL -550
LQGYNLSPKVDPQLNERNPFRSNP -574
```

B

```
MGTSSGSNHPHQMLPPRQQQRSGGGLETALSLVSSDQEPRRESPAESASS -50
QETWPLGDTVAGKKSMSQKTEPDSMEQTVNVMHHVSNADKVSVRDIARER -100
VELVAERMHRLPDEFLDELKNGLKSILEGNVAQSVDEFMFLQKVVQSRTD -150
LSSVTLVRAHRVQLEILVAINTGIQAFLHPNISLSQPSLIEIFVYKRCRN -200
IACQNQLPADDCYCDICTNRKGFCNLCMCTICNKFDFSVNTCRWIGCDLC -250
SHWTHTDCAIRDGQITTGSSAKNNTSGPGEIVFKCRACNRTSELLGWVKD -300
VFQHCAPNWDRESLMKELDFVSRIFRGSEDQRGRKLFWKCEELIDKIKGG -350
LAEATAAKLILMFFQEIESDSAKSFENGEGGRLMAPQDACNRIAEVVQET -400
LRKMEIVAEEKMRMFKKARMALETCDRELEDKAKEVSELKAERQKKKLQI -450
DELERIVRLKQAEADMFQLKANEAKREADRLQRIVLAKMDKSEEEYASNY -500
LKQRLSEAEAEKQYLFEKIKLQENSRVASQSSGGGGDPSQVMMYSKIRDL -550
LQGYNLSPKVDPQLNERNPFRSNP -574
```

FIGURE 8

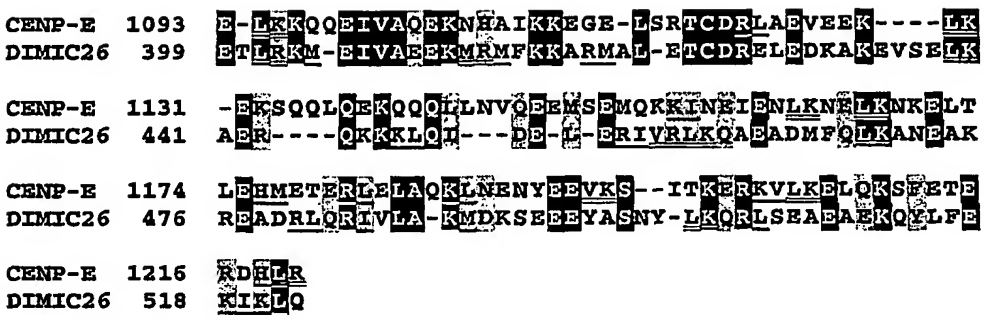
CENP-E 1093 E-LKKQQEIVAQEKNHAIKKEGE-LSRTCDRLAEVEEK----LK
DIMIC26 399 ETLRKM-EIVAEEKMRMFKKARMAL-ETCDRELEDKAKEVSELK
CENP-E 1131 -EKCSQQLQEKQQQLLNVQEEMSEMQKKINEIENLKNELKNKELT
DIMIC26 441 AER----QKKKLQI---DE-L-ERIVRLKQAEADMFQLKANEAK
CENP-E 1174 LEHMETERLELAQKLNENYEEVKS--ITKERKVLKELQKSEETE
DIMIC26 476 READRLQRIVLA-KMDKSEEEYASNY-LKQRLSEAEAEKQYLFE
CENP-E 1216 RDHLR
DIMIC26 518 KIKLQ

FIGURE 9

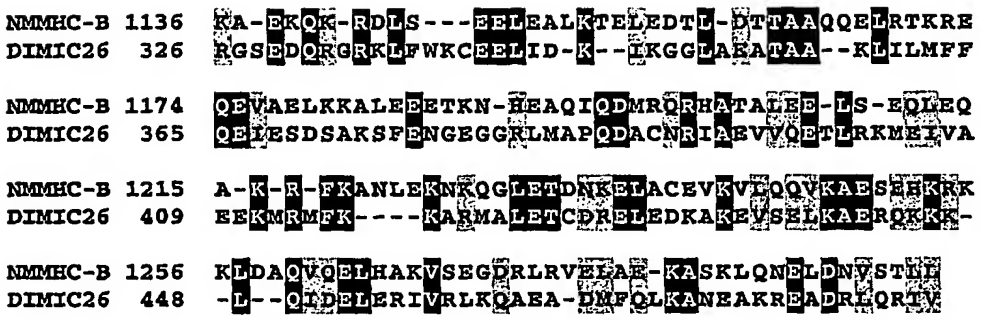
NMMHC-B 1136 KA-EKQK-RDLS---EELEALKTELEDTL-DTTAAQQELRTKRE
DIMIC26 326 RGSEDQRGRKLFWKCEELID-K--IKGGLAEATAA--KLILMFF
NMMHC-B 1174 QEVAELKKALEEETKN-HEAQIQDMRQRHATALEE-LS-EQLEQ
DIMIC26 365 QEIESDSAKSFENGEGGRLMAPQDACNRIAEVVQETLRKMEIVA
NMMHC-B 1215 A-K-R-FKANLEKNKQGLETDNKELACEVKVLQQVKAESEHKKK
DIMIC26 409 EEKMRMFK----KARMALETCDRELEDKAKEVSELKAERQKKK-
NMMHC-B 1256 KLDAQVQELHAKVSEGDRLRVELAE-KASKLQNELDNVSTLI
DIMIC26 448 -L--QIDELERIVRLKQAEA-DMFQLKANEAKREADRLQRTV

FIGURE 10

```
Motif 1 DIMIC26/1
      212 -CYCDIC- 217
      227 -CMCTIC- 232
consensus-CXCXIC

Motif 2 DIMIC26/2
      287 -ACNRTSELL- 295
      389 -ACNRIAEVV- 397
consensus-ACNRXXE[M/I/L/V][M/I/L/V]

Motif 3 DIMIC26/3
       20 -QRSGGG- 25
      530 -QSSGGG- 535
consensus-QXSGGG

Motif 4 DIMIC26/4
      116 -LDELKNGL- 123
      344 -IDKIKGGL- 351
consensus-[M/I/L/V]DX[M/I/L/V]KXGL

Motif 5 DIMIC26/5
      437 -SELKAEKQ- 444
      506 -SEAEAEKQ- 513
consensus-SEXXAEKQ

Motif 6 DIMIC26/6
      458 -RLKQAEAD- 465
      504 -RLSEAEAE- 511
consensus-RLXXAEA[D/E]
```

FIGURE 11

A

```
65587                                                                atggctg
65580 acaaagtcga taccaagttc gtctgggttc cgatttccgt cccaagtagg acggttcaag
65520 cgcatctaca gaaaaacggt ttgtactttc aaaagaacgc aacttttatt tcgaaagttc
65460 gatcctttgg gtccatagta ttgacgggtg agtagattac aattctgatt gctttagggt
65400 ttttgccgat ttggggatgg ttgatctttc agttgaatcc ttaatctgat ttgctgacat
65340 tagaaggttc attccatgtt ttgtcttaat ccccatctag ttgccgaaag ggtttagatt
65280 ttggggtcat tgcttgaaat ttatgcttca gatgcgttga tcttgtaccg tttagaactc
65220 tgtgttgtac tcggattcaa agttgttccc tttacatgtc atatcttctgt ttttgtttatt
65160 ttcatgcctc accattgaat tcttcgtttc aggccactca accaactacg agtgctatct
65100 ttcacccctg agttcactga aagtgttcct gcttggtttg ctgatgagct attgcttgtg
65040 gagttagttt aaatatctgc aaatgttcaag tcttcttaat gtactgccac atatgcttta
64980 tatcactatt agctatgtaa cttctaggat tatttcaatg caaagatgac ccccctgctt
64920 tggttctgct tcaaaatttt gatttttgat tctaatttgc agatatctca gtgtagctat
64860 catggatact caggactcac tagtagggga ggcataaaca ctgttgaaaa ccatagatgg
64800 gtctggcaca acaacggagt caggttgtcg tttccaagag ccgagtcttc tataaacatc
64740 actatgggtt gtacgcttca gcgtgggata gcaaaaagct taagtcagga aaacctagtg
64680 gagttatctg atgaaaatga tgatctatgt cctgtggagt gtgtcactga gttcaagaca
64620 gatgatgaat tgcttagcgt tcttgaaaag tcgaaagaaa ctaattcttt ggttgtggtt
64560 gatttttatc gcactgcatg tgggagttgt aaatacatag agcagggctt ctcaaaactg
64500 tgcaagcaat ctggttgacca agaagctcct gttatcttcc ttaagcataa tgtgagtttc
64440 tgatctctca tctttgatgt ataaacaata aagaagtaga gaaaaaatcc attaatagcg
64380 ctaatgctgc tatttgcagg tggtagatga atatgatgaa caatctgaag tcgcagaaag
64320 gctccgtatc aaggtaatcc ctctttagaa tttcggtact aacaacactc aaaaattaag
64260 tgaacagcgt agtaaagagt ttgatatact acaatcttac aggcggttcc tctcttccac
64200 ttctacaaaa acggagttct cttagaatca tttgcaacta gagacaagga gaggatcgac
64140 gcagctattc tcaaatatac atcctcggaa tcttga
64104 agaa cattcgacaa aaccgcactc
64080 tttgggttac ctttcccaaa tcattgaata tgatgtaact tatattgaac aacacagcat
64020 catccttctt cttgtaagat gtcatactga ataagacaga tgcaacattt tgatctc [a
      aaaaaaaaaa aaaaaaaaa]
```

B

```
65587                                                                atggctg
65580 acaaagtcga taccaagttc gtctgggttc cgatttccgt cccaagtagg acggttcaag
65520 cgcatctaca gaaaaacggt ttgtactttc aaaagaacgc aacttttatt tcgaaagttc
65460 gatcctttgg gtccatagta ttgacgggtg agtagattac aattctgatt gctttagggt
65400 ttttgccgat ttggggatgg ttgatctttc agttgaatcc ttaatctgat ttgctgacat
65340 tagaaggttc attccatgtt ttgtcttaat gcccatctag ttgccgaaag ggtttagatt
65280 ttggggtcat tgcttgaaat ttatgcttca gatgcgttga tcttgtaccg tttagaactc
65220 tgtgttgtac tcggattcaa agttgttccc tttacatgtc atatcttctgt ttttgtttatt
65160 ttcatgcctc accattgaat tcttcgtttc aggccactca accaactacg agtgctatct
65100 ttcacccctg agttcactga aagtgttcct gcttggtttg ctgatgagct attgcttgtg
65040 gagttagttt aaatatctgc aaatgttcaag tcttcttaat gtactgccac atatgcttta
64980 tatcactatt agctatgtaa cttctaggat tatttcaatg caaagatgac ccccctgctt
64920 tggttctgct tcaaaatttt gatttttgat tctaatttgc agatatctca ctgtagctat
64860 catggatact caggactcac tagtagggga ggcataaaca ctgttgaaaa ccatagatgg
64800 gtctggcaca acaacggagt caggttgtcg tttccaagag ccgagtcttc tataaacatc
64740 actatgggtt gtacgcttca gcgtgggata gcaaaaagct taagtcagga aaacctagtg
64680 gagttatctg atgaaaatga tgatctatgt cctgtggagt gtgtcactga gttcaagaca
64620 gatgatgaat tgcttagcgt tcttgaaaag tcgaaagaaa ctaattcttt ggttgtggtt
64560 gatttttatc gcactgcatg tgggagttgt aaatacatag agcagggctt ctcaaaactg
64500 tgcaagcaat ctggttgacca agaagctcct gttatcttcc ttaagcataa tgtgagtttc
64440 tgatctctca tctttgatgt ataaacaata aagaagtaga gaaaaaatcc attaatagcg
64380 ctaatgctgc tatttgcagg tggtagatga atatgatgaa caatctgaag tcgcagaaag
64320 gctccgtatc aaggtaatcc ctctttagaa tttcggtact aacaacactc aaaaattaag
64260 tgaacagcgt agtaaagagt ttgatatact acaatcttac aggcggttcc tctcttccac
64200 ttctacaaaa acggagttct cttagaatca tttgcaacta gagacaagga gaggatcgac
64140 gcagctattc tcaaatatac atcctcggaa tcttga
64104 agaa cattcgacaa aaccgcactc
64080 tttgggttac ctttcccaaa tcattgaata tgatgtaact tatattgaac aacacagcat
64020 catccttctt cttgtaagat gtcatactga ataagacaga tgcaacattt tgatctc [a
      aaaaaaaaaa aaaaaaaaa]
```

FIGURE 12

A

```
1                                                              atggctg
8    acaaagtcga taccaagttc gtctgggttc cgatttccgt cccaagtagg acggttcaag
68   cgcatctaca gaaaaacggt ttgtactttc aaaagaacgc aacttttatt tcgaaagttc
128  gatcctttgg gtccatagta ttgacgggcc actcaaccaa ctacgagtgc tatctttcac
188  ccctgagttc actgaaagtg ttcctgcttg gtttgctgat gagctattgc ttgtggagac
248  tcactagtag gggaggcata aacactgttg aaaaccatag atgggtctgg cacaacaacg
308  gagtcaggtt gtcgtttcca agagccgagt cttctataaa catcactatg ggttgtacgc
368  ttcagcgtgg gatagcaaaa agcttaagtc aggaaaacct agtggagtta tctgatgaaa
428  atgatgatct atgtcctgtg gagtgtgtca ctgagttcaa gacagatgat gaattgctta
488  gcgttcttga aaagtcgaaa gaaactaatt ctttggttgt ggttgatttt tatcgcactg
548  catgtgggag ttgtaaatac atagagcagg gcttctcaaa actgtgcaag caatctggcc
608  aagaagctcc tgttatcttc cttaagcata atgtggtaga tgaatatgat gaacaatctg
668  aagtcgcaga aaggctccgt atcaaggcgg ttcctctctt ccacttctac aaaaacggag
728  ttctcttaga atcatttgca actagagaca aggagaggat cgacgcagct attctcaaat
788  atacatcctc ggaatcttga agaacattcg acaaaaccgc actctttggg ttacctttcc
848  caaatcattg aatatgatgt aacttatatt gaacaacaca gcatcatcct tcttcttgta
908  agatgtcata ctgaataaga cagatgcaac attttgatct caaaaaaaaa aaaaaaaaaa
968  a
```

B

```
MADKVDTKFVWVPISVPSRTVQAHLQKNGLYFQKNATFISKVRSFGSIVL   -50
TGHSTNYECYLSPLSSLKVFLLGLLMSYCLWRLTSRGGINTVENHRWVWH   -100
NNGVRLSFPRAESSINITMGCTLQRGIAKSLSQENLVELSDENDDLCPVE   -150
CVTEFKTDDELLSVLEKSKETNSLVVV[DFYRTACGSCKYIEQGFSKLCKQ  -200
SGQEAPVIFLKHNV]VDEYDEQSEVAERLRIKAVPLFHFYKNGVLLESFAT  -250
RDKERIDAAILKYTSSES                                    -268
```

FIGURE 13

A

```
1                                                              atggctg
8   acaaagtcga taccaagttc gtctgggttc cgatttccgt cccaagtagg acggttcaag
68  cgcatctaca gaaaaacggt ttgtactttc aaaagaacgc aacttttatt tcgaaagttc
128 gatcctttgg gtccatagta ttgacgggcc actcaaccaa ctacgagtgc tatctttcac
188 ccctgagttc actgaaagtg ttcctgcttg gtttgctgat gagctattgc ttgtggagac
248 tcactagtag gggaggcata aacactgttg aaaaccatag atgggtctgg cacaacaacg
308 gagtcaggtt gtcgtttcca agagccgagt cttctataaa catcactatg ggttgtacgc
368 ttcagcgtgg gatagcaaaa agcttaagtc aggaaaacct agtggagtta tctgatgaaa
428 atgatgatct atgtcctgtg gagtgtgtca ctgagttcaa gacagatgat gaattgctta
488 gcgttcttga aaagtcgaaa gaaactaatt ctttggttgt ggttgatttt tatcgcactg
548 catgtgggag ttgtaaatac atagagcagg gcttctcaaa actgtgcaag caatctggtg
608 accaagaagc tcctgttatc ttccttaagc ataatgtggt agatgaatat gatgaacaat
668 ctgaagtcgc agaaaggctc cgtatcaagg cggttcctct cttccacttc tacaaaaacg
728 gagttctctt agaatcattt gcaactagag acaaggagag gatcgacgca gctattctca
788 aatatacatc ctcggaatct tga
```

B

```
MADKVDTKFVWVPISVPSRTVQAHLQKNGLYFQKNATFISKVRSFGSIVL   -50
TGHSTNYECYLSPLSSLKVFLLGLLMSYCLWRLTSRGGINTVENHRWVWH   -100
NNGVRLSFPRAESSINITMGCTLQRGIAKSLSQENLVELSDENDDLCPVE   -150
CVTEFKTDDELLSVLEKSKETNSLVVV[DFYRTACGSCKYIEQGFSKLCKQ  -200
SGDQEAPVIFLKHNV]VDEYDEQSEVAERLRIKAVPLFHFYKNGVLLESFA  -250
TRDKERIDAAILKYTSSES                                   -269
```

FIGURE 14

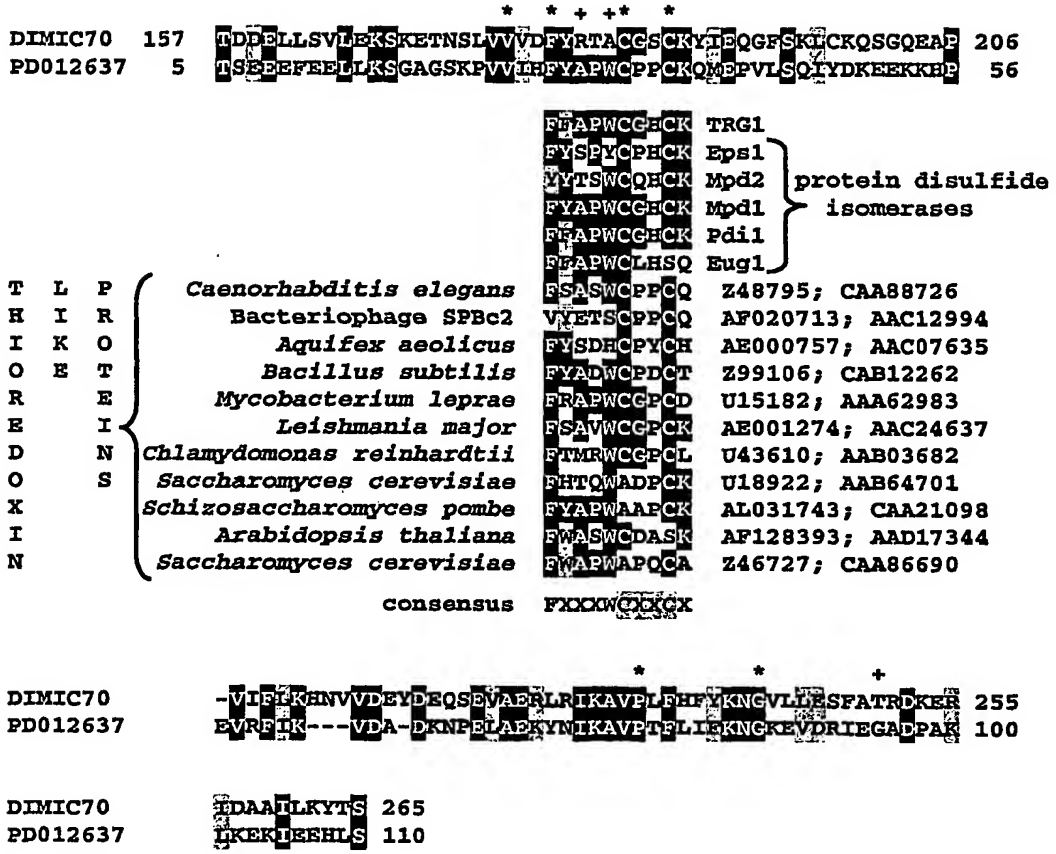
DIMIC70 157 TDDELLSVLEKSKETNSLVVVDFYRTACGSCKYIEQGFSKLCKQSGQEAP 206
PD012637 5 TSEEEFEELLKSGAGSKPVVIHFYAPWCPPCKQMEPVLSQIYDKEEKKEP 56
FEAPWCGHCK TRG1
FYSPYCPHCK Eps1
MYTSWCQHCK Mpd2
FYAPWCGHCK Mpd1
FEAPWCGHCK Pdi1
FEAPWCLHSQ Eug1
protein disulfide isomerases
T L P Caenorhabditis elegans FSASWCPPCQ Z48795; CAA88726
H I R Bacteriophage SPBc2 VYETSCPPCQ AF020713; AAC12994
I K O Aquifex aeolicus FYSDHCPYCH AE000757; AAC07635
O E T Bacillus subtilis FYADWCPDCT Z99106; CAB12262
R E Mycobacterium leprae FRAPWCGPCD U15182; AAA62983
E I Leishmania major FSAVWCGPCK AE001274; AAC24637
D N Chlamydomonas reinhardtii FTMRWCGPCL U43610; AAB03682
O S Saccharomyces cerevisiae FHTQWADPCK U18922; AAB64701
X Schizosaccharomyces pombe FYAPWAAPCK AL031743; CAA21098
I Arabidopsis thaliana FWASWCDASK AF128393; AAD17344
N Saccharomyces cerevisiae FWAPWAPQCA Z46727; CAA86690
consensus FXXXWCXXCX
DIMIC70 -VIFLKHNVVDEYDEQSEVAERLRIKAVPLFHFYKNGVLLESFATRDKER 255
PD012637 EVRFIK---VDA-DKNPELAEKYNIKAVPTFLIEKNGKEVDRIEGADPAK 100
DIMIC70 IDAAILKYTS 265
PD012637 LKEKIEEHLS 110

FIGURE 15

A

```
   1  ATGGCTGACA AAGTCGATAC CAAGTTCGTC TGGGTTCCGA TTTCCGTCCC
  51  AAGTAGGACG GTTCAAGCGC ATCTACAGAA AAACGGTTTG TACTTTCAAA
 101  AGAACGCAAC TTTTATTTCG AAAGTTCGAT CCTTTGGGTC CATAGTATTG
 151  ACGGGCCACT CAACCAACTA CGAGTGCTAT CTTTCACCCC TGAGTTCACT
 201  GAAAGTGTTC CTGCTTGGTT TGCTGATGAG CTATTGCTTG TGGAGTTAGT
 251  TTAAATATCT GCAAATGTCA AGTCTTCTTA ATATATCTCA CTGTAGCTAT
 301  CATGGATACT CAGGACTCAC TAGTAGGGGA GGCATAAACA CTGTTGAAAA
 351  CCATAGATGG GTCTGGCACA ACAACGGAGT CAGGTTGTCG TTTCCAAGAG
 401  CCGAGTCTTC TATAAACATC ACTATGGGTT GTACGCTTCA GCGTGGGATA
 451  GCAAAAAGCT TAAGTCAGGA AAACCTAGTG GAGTTATCTG ATGAAAATGA
 501  TGATCTATGT CCTGTGGAGT GTGTCACTGA GTTCAAGACA GATGATGAAT
 551  TGCTTAGCGT TCTTGAAAAG TCGAAAGAAA CTAATTCTTT GGTTGTGGTT
 601  GATTTTTATC GCACTGCATG TGGGAGTTGT AAATACATAG AGCAGGGCTT
 651  CTCAAAACTG TGCAAGCAAT CTGGTGACCA AGAAGCTCCT GTTATCTTCC
 701  TTAAGCATAA TGTGGTAGAT GAATATGATG AACAATCTGA AGTCGCAGAA
 751  AGGCTCCGTA TCAAGGCGGT TCCTCTCTTC CACTTCTACA AAAACGGAGT
 801  TCTCTTAGAA TCATTTGCAA CTAGAGACAA GGAGAGGATC GACGCAGCTA
 851  TTCTCAAATA TACATCCTCG GAATCTTGAA
```

B

```
   1  MSSLLNISHC SYHGYSGLTS RGGINTVENH RWVWHNNGVR LSFPRAESSI
  51  NITMGCTLQR GIAKSLSQEN LVELSDENDD LCPVECVTEF KTDDELLSVL
 101  EKSKETNSLV VVDFYRTACG SCKYIEQGFS KLCKQSGDQE APVIFLKHNV
 151  VDEYDEQSEV AERLRIKAVP LFHFYKNGVL LESFATRDKE RIDAAILKYT
 201  SSES
```

Figure 16

crop025pct.ST25.txt
SEQUENCE LISTING

<110> CROPDESIGN N.V.

<120> Nucleic acid molecules encoding DIM interactors and uses therefor

<130> CROP-025-PCT

<150> US 60/259,890
<151> 2000-01-05

<160> 111

<170> PatentIn version 3.1

<210> 1
<211> 713
<212> DNA
<213> Arabidopsis thaliana

<400> 1
caaaaaaaga gatcgggaaa atcgtcgaga gagagagaga gagagagaat gtcgtatctt 60

cttccacatc tgcactccgg ttgggctgtt gatcagtcga ttctggccga ggaagagcgt 120

ctcgtcgtca ttcgtttcgg ccatgactgg gatgagacct gtatgcagat ggatgaggtg 180

cttgcgtctg ttgctgagac gattaagaac tttgcagtca tttatctggt ggacatcact 240

gaggttccag acttcaacac catgtacgag ctgtacgatc cttctacggt catgttcttc 300

ttcaggaaca agcacatcat gatcgatctt ggaactggta acaacaacaa gatcaactgg 360

gctctcaagg acaagcagga gttcattgat atcattgaga ctgtctaccg tggtgcaagg 420

aagggtcgtg ggttggtgat tgctccaaaa gattactcca ccaaataccg ttactaatcg 480

agcttcccaa cactatctag tttgttaaaa ccattgagtc tagtgattct ggtcagctga 540

aatatcccgt gaactggttt catttcatat atgctttgat gatgattgtg attctggttt 600

ttgcttactg tttcggttgt tctccattta tgtatgtaac tacttgcttc gaacatgatt 660

cagtgaaatg aaatttggaa gccgttgggt ttatgtaaaa aaaaaaaaaa aaa 713

<210> 2
<211> 533
<212> DNA
<213> Glycine max

<400> 2
gaaattaggg ttgatcgagg aagaagaaga agatgtcgta cctgttgcct cacttgcact 60

ctgggtgggc ggttgatcaa gcaatcctgg cagaagagga gcgtctggtc gtgatccgtt 120

tcgggcacga ctgggacgaa acctgcatgc aaatggacga ggtgttggcg tcggtggcgg 180

agacgataaa gaactttgcg gtgatatacc tggttgacat aacggaggtg ccggatttca 240

acaccatgta cgagctctac gacccttcca cggtgatgtt cttcttcagg aacaagcaca 300

tcatgattga tctgggaacc ggtaacaaca acaagatcaa ctgggctctc aaggacaagc 360

aggagttcat cgacattgtt gagaccgtct accgtggagc caggaaggga cgtggtctcg 420

tcatcgctcc caaagattac tccaccaagt accgttacta gtttcattca tctgtaacat 480

acaacaaact aaatattatc tcttttctta cttctctggt gtggtgctgg ttg 533

<210> 3
<211> 602
<212> DNA
<213> Medicago truncatula

<220>
<221> misc_feature
<222> (378)..(378)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (503)..(503)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (533)..(533)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (542)..(542)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (544)..(544)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (560)..(560)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (579)..(579)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (584)..(585)
<223> N = any nucleic acid

<220>
<221> misc_feature

<222> (591)..(591)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (598)..(598)
<223> N = any nucleic acid

<400> 3
aagatgtcgt acctgctgcc tcacttgcat tcagggtggg ctgtggatca ggctatccta 60

gctgaggaag agcgtcttgt tgtcatccgt tttggtcatg attgggatga gacctgcatg 120

cagatggatg aggtgctgtc atcggttgca gagacaataa agaactttgc tgtgatatac 180

cttgtggaca tcacagaggt gcctgatttc aacactatgt atgagctgta cgacccttcg 240

acagtcatgt tcttctttag gaacaagcac atcatgatag atcttggaac tggaaacaac 300

aacaaaatca attgggctat gaaggacaaa caggagttca ttgacatcat tgaaactgtc 360

tacagaggtg caaggaangg acgtggtctc gtcattgctc ccaaagatta ctcaaccaag 420

tatcggtact agaatatata tctatctcta ttaggatgtg taatttgaac tctggatttt 480

caattctcat ccaaaactac ctnattgggt cactcttgac aatgtgaaac tgncactcaa 540

tntntcacca cgtggctggn tcttaaataa ctggctttnt gaannaatac nttatgcntc 600

ga 602

<210> 4
<211> 557
<212> DNA
<213> Lycopersicon esculentum

<400> 4
gtttggtgag cttaactcag aaagatgtcg tacttgttac ctcgccttca ctcaggatgg 60

gccgtcgatc aagccattct cgctgaagag gaacgccttg tcatcatccg cttcggccat 120

gactgggatg aaacttgtat gcagatggat gacgtgctag cttcagttgc agatacatta 180

aagaattttg ctgtgatata cctggtggac ataacagagg tccctgattt taacacgatg 240

tatgaattgt acgatccatc aactatcatg ttcttcttca ggaacaagca catcatgatt 300

gatcttggcg caggaaacaa caacaagatc aactgggcac tcaaagataa acaggagttc 360

attgatattg ttgagacagt gtaccgtggt gcacgaaagg gtcgtggtct agttattgca 420

ccaaaagatt actctaccaa gtatcgttac tgagtgtgtt gatattgtat tttgtttgtg 480

gaatatgctg cacttttatg actgaactgg aacatacagg tgctgagtgg tttccattag 540

actgattcaa ttgctgt 557

<210> 5
<211> 715

<212> DNA
<213> Gossypium arboreum

<400> 5
ggcacgaggt ttgatatcac aaattttgcg agaggaaaag catttcatcg tcagccaaaa 60

aaagggaatt ttagtcttca ggggaacgat cgaacgatgt cgtacctact accacacctg 120

cactcaggat gggccgtaga tcaggccatc ttggcagagg aagagcgtct cgtcgtcatc 180

cgattcggcc acgactggga cgaaacttgc atgcagatgg atgaggtgct tgcttctgtc 240

gctgaaacaa taaagaactt tgctgtgatt tatcttgtag acataacgga agtacctgat 300

tttaacacga tgtatgagct ttacgatcca tcaacagtga tgttcttttt cagaaacaag 360

cacattatga ttgatcttgg cactggtaat aacaacaaga ttaactgggc tctcaaggac 420

aaacaggagt tcatagacat cattgagact gtataccgtg gggcaaggaa aggtcgtggt 480

ctggtaattg ctccgaaaga ttactcgaca aaataccgct actgagtatt cttgcaacct 540

gttgttgatc agaaagaatt aatggtcata tttcgcgaat ttaatatgga ctattttcat 600

tatgaactgt gtatgtcatt tgcaatgctt gcatgaacgt ggtggataat cccagttggg 660

atgaaaactt gaaatatata atttatgggg tttttctaaa aaaaaaaaaa aaaaa 715

<210> 6
<211> 419
<212> DNA
<213> Lotus japonicus

<400> 6
cttaaccgag ttcgcatcgc atcgcaacgc agacctaata agagaagggt gcagatatgt 60

cgtacctgtt gcctcacttg cattcaggat gggctgtgga tcaggcaatc cttgctgagg 120

aagagcgtct tgttgtcatc cgcttcggtc acgattggga tgagacctgc atgcagatgg 180

atgaggtgct ggcatcggtt gcggagacaa taaaaaactt tgctgtaata taccttgtgg 240

acatcacgga ggtaccggat ttcaatacca tgtatgagct gtatgacccc tccactgtta 300

tgtttttctt cagaaacaag cacatcatga tagatcttgg cactggtaac aacaacaaga 360

tcaattgggc tatgaaggac aagcaggagt tcattgacat cattgagacc gtctacaga 419

<210> 7
<211> 626
<212> DNA
<213> Zea mays

<400> 7
aggggagca agcagcagcc gtccgtggaa gaagatgtcg tacctgctgc cgcatctgca 60

ctcggggtgg gcggtggacc aggccatcct cgccgaggag gagcgcctcg tcatcatccg 120

cttcggccat gactgggacg agacctgcat gcagatggat gaagtgctgg cagcagtagc 180

tgagaccata aagaactttg cggtcatcta ccttgttgac atcacagagg tccctgattt 240

caacaccatg tatgagctgt acgacccttc gacagtgatg ttcttcttcc ggaacaagca 300

catcatgatt gatctcggga ctggaaacaa caacaagatc aactgggccc tgaaggacaa 360

gcaagagttc atcgacattg tggagaccgt atacaggggc gcccggaaag gccgtggtct 420

ggtgattgct ccaaaggact actccaccaa gtaccgttac taagaaggca ggcattcccc 480

ttatgcaatc tattggggat gtttacacgg acagttagtc cgacaggtga tccctaactg 540

tatcttggca taccctagta attcaagttg ctggctggag tgatgcttat ccatgtcttg 600

gcatgccata gtgaacatgt aaacct 626

<210> 8
<211> 721
<212> DNA
<213> Oryza sativa

<400> 8
gcgagtcgcc gaagcggagg aagaagcagc agcttgcggc ggcggcggcg aggcggagat 60

gtcgtacctt ctgccgcatc tgcactcggg gtgggcggtg gaccaggcca tcctcgcgga 120

ggaggagcgc ctcgtcatca tccgcttcgg ccacgactgg gacgagacgt gcatgcagat 180

ggatgaggtg ctagcggcag tagctgagac cataaagaac tttgcggtca tctacctggt 240

cgacatcacc gaggttccag acttcaacac catgtacgag ctgtatgatc cgtcgacggt 300

gatgttcttc ttccgcaaca agcacatcat gattgatctg gggacaggaa acaacaacaa 360

gatcaactgg gccttgaagg acaagcagga gttcatcgac attgtggaga ccgtctacag 420

aggtgctcgc aaagggcgtg gtttggtgat tgctcccaag gactactcca ctaaataccg 480

ttactaagct atattttgcc ccacgctgcc gttgagtggc ccttagtaca agagctgtaa 540

gcttatggag ttgtggtggg gtgttgttga ctaaaaactg taatattgtt cgaaactcga 600

gtgtcggaac ataatttaca tgtcaattgg tacatcaatg acaagttgat tggtgtggat 660

accaatcatc cattaattat caatactctt tccagtctta taaaaaaaaa aaaaaaaaaa 720

a 721

<210> 9
<211> 484
<212> DNA
<213> Hybrid aspen

<400> 9
gagagagaga tgtcgtacct gctacctcac ttgcactcag gatgggctgt agaccaggcg 60

attctagctg aagaagagcg attggtagtc attcgatttg gtcatgattg ggatgagact 120

tgtatgcaga tggacgaagt gttggcatca gtagcagaaa cgataaagaa ttttgctgtg 180

atatacctgg tggacatcac agaggttcct gattttaaca caatgtatga gttgtatgat 240

ccatctactg tcatgttctt cttcaggaac aagcatatta tgattgatct tggaactggt 300

aataacaaca agatcaactg ggctctcaaa gacaagcagg aattcattga cattgttgag 360

actgtctacc ggggggcaag gaagggtcgt ggtcttgtca ttgcaccgaa agactactcc 420

acaaagtacc gctactaagt acttttgcta attgtttgaa ctgtagtata ttgccttgaa 480

aaca 484

<210> 10
<211> 514
<212> DNA
<213> Pinus taeda

<220>
<221> misc_feature
<222> (9)..(11)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (448)..(448)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (450)..(451)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (454)..(455)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (502)..(508)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (511)..(512)
<223> N = any nucleic acid

<400> 10
gatcgagcnn nggcaaaaat gtcgtacttg ctaccccatt tgcactccgg gtgggcagtg 60

gatcaggcca ttctagccga ggaggagcgt ctcgtcgtca ttcgcttcgg ccatgactgg 120

gacgaaacat gtatgcagat ggacgaagtg ctttcatctg ttgcagaaac tatcaagaat 180

ttcgcagtga tatacctagt ggacataacg gaggtgcctg atttcaacac catgtacgaa 240

ctgtatgacc cgtctactgt catgtttttc tttcgcaata aacatataat gatagatttg 300

gggacgggga acaacaataa aatcaactgg gctatgaagg acaagcagga attcattgat 360

atcattgaga ccgtgtacag aggggccagg aagggcaggg gtctggtcat tgcccccaag 420

gattattcta ccaagtaccg atactagntn natnncagaa attgtattct atttcatatg 480

ctgaactggg ttttctttaa cnnnnnnnat nnat 514

<210> 11
<211> 618
<212> DNA
<213> Triticum aestivum

<400> 11
cacctcggct gttcgtctcc acattattct ctccggcggc ggcgactcga ctctcggcga 60

ggggcggaga tgtcctacct actcccgcac ctgcactccg gctgggcggt ggaccaggcc 120

atcctcgccg aggaggagcg gctggtcatg atccgcttcg gccacgactg ggacgagacg 180

tgcatgcaga tggacgaggt gctgtcaggg gtggctgaga ccataaagaa ctttgcggtg 240

atctacctcg tcgacatcac ggaggttcct gacttcaaca ccatgtacga gctgtatgat 300

ccgtcaacgg tgatgttctt cttccgcaac aagcacatca tgattgatct cgggacggga 360

aacaacaaca agatcaactg ggcaatgaaa gacaagcaag agtttgttga cattgtggag 420

actgtctaca gaggagctcg taaggggcgt ggtctggtga ttgctcccaa ggattactcc 480

acaaaatacc gttactgagg tatctgtgcc tggctatcca tgtgcagtgt gtggcctgct 540

ctgttccgtg atatgcgctt tgtgaccctt gatgtaatgt ctgtcaactt tgaaagtgtc 600

gtaactatgt atttgtat 618

<210> 12
<211> 785
<212> DNA
<213> Hordeum vulgare

<400> 12
ccgttatgat gctctataga ctactcgggg gactgagctg gagctccacc gcggaggcgg 60

acgctctaga actagtggat cccccgggct gcaggaattc ggcacgtgcg gcggcgacgg 120

atctcggcga gatcggagat gtcttacctg ctcccgcacc ttcactcggg ctgggcggtg 180

gaccaggcca tcctcgccga ggaggagcgg ctggtcatga tccgcttcgg ccacgactgg 240

gacgagacgt gcatgcagat ggacgaggtg ctgtcagggg tggctgagac aataaagaac 300

tttgcggtga tctaccttgt cgacatcacc gaggtccccg acttcaacac catgtacgag 360

ctgtacgacc cgtcgacggt catgttcttc ttccgcaaca agcacatcat gattgatctc 420

gggacaggaa acaacaacaa gatcaactgg gcaatgaaag acaagcaaga gtttgttgac 480

atcgtggaga ctgtctacag gggagctcgt aagggccgtg gtctggtgat tgctcccaag 540

gattactcca ccaaataccg ttactgaggt atctgtgcct ggctatatgg cctactgtga 600

aatgcctgtt ctcataaacc tataactcag gctgtgaccc tttatgtaat gtttttcaac 660

tttgaaagtg ctgtgcaact tggatttgta tcttgaatat cgacgtgtca taaacctgta 720

acccttttgaa acatgaccta ctggctattg gtatgactgg gaaaacagac ctttggtttg 780

aaaaa 785

<210> 13
<211> 239
<212> DNA
<213> Thellungiella salsuginea

<220>
<221> misc_feature
<222> (205)..(205)
<223> N = any nucleic acid

<400> 13
aatcgtcgag agagacaaga gaagatgtcg tatcttctac cacacttgca ctccggttgg 60

gctgtcgatc aggcgattct ggctgaggaa gagcgtctcg tcgtcatacg atttggccat 120

gattgggatg agacttgtat gcagatggat gaggtactgg catctgttgc tgagacgatc 180

aagaactctg cagtggatat ctggnggaca taacagaggt tcctgacttc aacaccatg 239

<210> 14
<211> 491
<212> DNA
<213> Cryptomeria japonica

<220>
<221> misc_feature
<222> (398)..(398)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (445)..(445)
<223> N = any nucleic acid

<220>
<221> misc_feature
<222> (449)..(449)
<223> N = any nucleic acid

<400> 14
caaaaaggag aagagcaagg gcagaaatgt cttacttgtt gccgcacctg cactcgggat 60

gggcggtgga ccaggcgata ctggcggagg aggagcggct agtggtaatt aggttcggcc 120

atgactggga tgacacctgc atgcagatgg acgaggtgct ttcgagcgtg gcagagtcca 180

tcaagaattt tgcggtgata tacctggtgg acatcacgga ggtgcccgac ttcaatacta 240

tgtatgagct ttacgaccct tccactgtga tgtttttctt ccgcaacaag cacataatga 300

ttgatcttgg gaccggcaac aacaacarga tcaactgggc tctcaaggac aagcaggagt 360

tcatcgacat tatcgaaacc gtgtatcgcg gtgcccgnaa gggccgcggt ctcgtcattg 420

cccccaagga cwattccacc aagantcgnw ctagttttgc attctctttc tattcgsata 480

ttacctacac a 491

<210> 15
<211> 421
<212> DNA
<213> Mesembryanthemum crystallinum

<400> 15
cataacagag gtgcctgatt tcaaacaatg tacgagctgt atgatccatc cacagttatg 60

ttctttttca gaaacaagca catcatgata gatcttggaa cggggaacaa caataagatc 120

aactgggccc tcaaggacaa gcaggagttc atagacatcg ttgagactgt ctatcgtgga 180

gcacggaaag gtcgaggtct tgtgattgca ccgaaagatt actccactaa gtatcgctat 240

taagtcattg ttgttgctaa tctatcagat tgcaactcga gaagtgtaat attaagttga 300

taagccctgc atctaggatt cgtaaggttc tcttgaattc atgtgacagt gagtgatgta 360

ctttctattg atgttaactg tacccacaat gagaaattga gaaggtaagt tcctcacctt 420

t 421

<210> 16
<211> 432
<212> DNA
<213> Schizosaccharomyces pombe

<400> 16
atgagttatt ttttacctca tttacattct ggatggcacg ttgatcaagc aattttatca 60

gaacaagagc gcttggttgt cattcgattt ggtagagatc atgatgaaga atgtataaaa 120

caagatgaag tcctatacag aattgctgaa aaggtcgtca acatggctgt catatattta 180

gtggacattg acgaagttcc tgacttcaac aagatgatgt acgagcttta tgacagaact 240

acaataatgt ttttttatcg aaacaaacac atgatgattg acttaggtac tggtaacaat 300

aacaaaatca actggccgtt agaagataaa caagagatga tcgatattat agaaactatt 360

tttcgtggtg ctagaaaagg taaaggtctt gttatttctc caaaagatta ctctacccga 420

catcgctact ag 432

<210> 17
<211> 429
<212> DNA
<213> Drosophila melanogaster

<400> 17

atgtcgtata tgctccctca tttgcacaat ggctggcagg tggaccaggc cattctctcc 60

gaggaggacc gagtagttgt aatacgtttc ggtcacgatt gggaccccgc ctgcatgaaa 120

atggatgagg tcatgtacag catcgccgag aaggtgaaga actttgctgt catctatttg 180

gtggacatta ccgaggtgcc ggacttcaac aagatgtacg agttgtacga tccttgcacg 240

gtgatgttct tcttccgcaa caagcacatc atgatcgatt tgggcacggg caacaacaac 300

aagatcaact ggccactgga ggacaagcag gagatgatcg acattgtgga aacggtgtat 360

cgaggtgccc gtaagggccg tggtctggta gtctcgccca aggactactc taccaagtac 420

agatactaa 429

<210> 18
<211> 142
<212> PRT
<213> Arabidopsis thaliana

<400> 18

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ser Ile Leu Ala Glu Glu Glu Arg Leu Val Val Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ala Ser Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Leu Lys Asp Lys Gln Glu Phe
100 105 110

Ile Asp Ile Ile Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 19
<211> 142
<212> PRT

<213> Glycine max

<400> 19

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Val Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ala Ser Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Leu Lys Asp Lys Gln Glu Phe
100 105 110

Ile Asp Ile Val Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 20
<211> 142
<212> PRT
<213> Medicago truncatula

<220>
<221> MISC_FEATURE
<222> (125)..(125)
<223> X = any amino acid

<400> 20

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Val Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ser Ser Val
35 40 45

```
Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
    50                  55                  60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65                  70                  75                  80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
                85                  90                  95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Met Lys Asp Lys Gln Glu Phe
            100                 105                 110

Ile Asp Ile Ile Glu Thr Val Tyr Arg Gly Ala Arg Xaa Gly Arg Gly
        115                 120                 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
    130                 135                 140
```

<210> 21
<211> 142
<212> PRT
<213> Lycopersicon esculentum

<400> 21

```
Met Ser Tyr Leu Leu Pro Arg Leu His Ser Gly Trp Ala Val Asp Gln
1               5                   10                  15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Ile Ile Arg Phe Gly His
            20                  25                  30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Asp Val Leu Ala Ser Val
        35                  40                  45

Ala Asp Thr Leu Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
    50                  55                  60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65                  70                  75                  80

Ile Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Ala
                85                  90                  95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Leu Lys Asp Lys Gln Glu Phe
            100                 105                 110

Ile Asp Ile Val Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
        115                 120                 125
```

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 22
<211> 142
<212> PRT
<213> Gossypium arboreum

<400> 22

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Val Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ala Ser Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Leu Lys Asp Lys Gln Glu Phe
100 105 110

Ile Asp Ile Ile Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 23
<211> 121
<212> PRT
<213> Lotus japonicus

<400> 23

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Val Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ala Ser Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Met Lys Asp Lys Gln Glu Phe
100 105 110

Ile Asp Ile Ile Glu Thr Val Tyr Arg
115 120

<210> 24
<211> 142
<212> PRT
<213> Zea mays

<400> 24

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Ile Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ala Ala Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Leu Lys Asp Lys Gln Glu Phe
100 105 110

Ile Asp Ile Val Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 25
<211> 142
<212> PRT
<213> Oryza sativa

<400> 25

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Ile Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ala Ala Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Leu Lys Asp Lys Gln Glu Phe
100 105 110

Ile Asp Ile Val Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 26
<211> 142
<212> PRT
<213> Hybrid aspen

<400> 26

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Val Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ala Ser Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Leu Lys Asp Lys Gln Glu Phe
100 105 110

Ile Asp Ile Val Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 27
<211> 142
<212> PRT
<213> Pinus taeda

<400> 27

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Val Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ser Ser Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Met Lys Asp Lys Gln Glu Phe
100 105 110

Ile Asp Ile Ile Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 28
<211> 142
<212> PRT
<213> Triticum aestivum

<400> 28

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Met Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ser Gly Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Met Lys Asp Lys Gln Glu Phe
100 105 110

Val Asp Ile Val Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 29
<211> 142
<212> PRT
<213> Hordeum vulgare

<400> 29

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Met Ile Arg Phe Gly His

20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ser Gly Val
35 40 45

Ala Glu Thr Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Ala Met Lys Asp Lys Gln Glu Phe
100 105 110

Val Asp Ile Val Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135 140

<210> 30
<211> 54
<212> PRT
<213> Thellungiella salsuginea

<400> 30

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Val Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Glu Thr Cys Met Gln Met Asp Glu Val Leu Ala Ser Val
35 40 45

Ala Glu Thr Ile Lys Asn
50

<210> 31
<211> 139
<212> PRT
<213> Cryptomeria japonica

<220>
<221> MISC_FEATURE
<222> (101)..(101)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (124)..(124)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (136)..(136)
<223> X = any amino acid

<400> 31

Met Ser Tyr Leu Leu Pro His Leu His Ser Gly Trp Ala Val Asp Gln
1 5 10 15

Ala Ile Leu Ala Glu Glu Glu Arg Leu Val Val Ile Arg Phe Gly His
20 25 30

Asp Trp Asp Asp Thr Cys Met Gln Met Asp Glu Val Leu Ser Ser Val
35 40 45

Ala Glu Ser Ile Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr
50 55 60

Glu Val Pro Asp Phe Asn Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr
65 70 75 80

Val Met Phe Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Xaa Ile Asn Trp Ala Leu Lys Asp Lys Gln Glu Phe
100 105 110

Ile Asp Ile Ile Glu Thr Val Tyr Arg Gly Ala Xaa Lys Gly Arg Gly
115 120 125

Leu Val Ile Ala Pro Lys Asp Xaa Ser Thr Lys
130 135

<210> 32
<211> 74
<212> PRT
<213> Mesembryanthemum crystallinum

<400> 32

Phe Gln Thr Met Tyr Glu Leu Tyr Asp Pro Ser Thr Val Met Phe Phe
1 5 10 15

Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr Gly Asn Asn Asn
20 25 30

Lys Ile Asn Trp Ala Leu Lys Asp Lys Gln Glu Phe Ile Asp Ile Val
35 40 45

Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly Leu Val Ile Ala
50 55 60

Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
65 70

<210> 33
<211> 142
<212> PRT
<213> Schizosaccharomyces pombe

<400> 33

Met Ser Tyr Phe Leu Pro His Leu His Ser Gly Trp His Val Asp Gln
1 5 10 15

Ala Ile Leu Ser Glu Gln Glu Arg Leu Val Val Ile Arg Phe Gly Arg
20 25 30

Asp His Asp Glu Glu Cys Ile Lys Gln Asp Glu Val Leu Tyr Arg Ile
35 40 45

Ala Glu Lys Val Val Asn Met Ala Val Ile Tyr Leu Val Asp Ile Asp
50 55 60

Glu Val Pro Asp Phe Asn Lys Met Tyr Glu Leu Tyr Asp Arg Thr Thr
65 70 75 80

Ile Met Phe Phe Tyr Arg Asn Lys His Met Met Ile Asp Leu Gly Thr
85 90 95

Gly Asn Asn Asn Lys Ile Asn Trp Pro Leu Glu Asp Lys Gln Glu Met
100 105 110

Ile Asp Ile Ile Glu Thr Ile Phe Arg Gly Ala Arg Lys Gly Lys Gly
115 120 125

Leu Val Ile Ser Pro Lys Asp Tyr Ser Thr Arg His Arg Tyr
130 135 140

<210> 34
<211> 139
<212> PRT
<213> Drosophila melanogaster

<400> 34

Met Leu Pro His Leu His Asn Gly Trp Gln Val Asp Gln Ala Ile Leu
1 5 10 15

Ser Glu Glu Asp Arg Val Val Val Ile Arg Phe Gly His Asp Trp Asp
20 25 30

Pro Ala Cys Met Lys Met Asp Glu Val Met Tyr Ser Ile Ala Glu Lys
35 40 45

Val Lys Asn Phe Ala Val Ile Tyr Leu Val Asp Ile Thr Glu Val Pro
50 55 60

Asp Phe Asn Lys Met Tyr Glu Leu Tyr Asp Pro Cys Thr Val Met Phe
65 70 75 80

Phe Phe Arg Asn Lys His Ile Met Ile Asp Leu Gly Thr Gly Asn Asn
85 90 95

Asn Lys Ile Asn Trp Pro Leu Glu Asp Lys Gln Glu Met Ile Asp Ile
100 105 110

Val Glu Thr Val Tyr Arg Gly Ala Arg Lys Gly Arg Gly Leu Val Val
115 120 125

Ser Pro Lys Asp Tyr Ser Thr Lys Tyr Arg Tyr
130 135

<210> 35
<211> 1348
<212> DNA
<213> Arabidopsis thaliana

<400> 35
ttcactgcaa atcatagagc agaagcagcg gcaaagcgtg gtgggagtgt gagtacatgc 60

ggagaaggta atgtggatgt tggaaacggt tatggtatac cgggtggagt tgcatacgcg 120

ggtcactccg agctcagtgg gaagcctgaa cctaccaatg cttcaaacaa tttgcctgaa 180

tatctcaaac aaaaattaaa ggccagggga attctcagag atggtgcagg ggctgttaca 240

tccaacccag aagatacatc agcggttagt tggaatagac aggcaacttt acctttcag 300

gctaatgcaa gcaccttacc attaggatgg gtagatgcaa aagatcctgc cagtggtgct 360

acctattact ataaccagca cactggaaca tgccagtggg aaaggcctgt tgagctttct 420

tatgccactt caagtgcacc acctgtactt tctaaagaag agtggattga aacatttgat 480

gaagcatctg gccataagta tttctacaat acaaggacac atgtgtctca gtgggaacct 540

ccagcttctt tacagaagcc cgctgccaca aactctaaca acgctgttac ccaaagtacg 600

gctaatggga agggggagca tcctccatct cagctgccaa gatgcagcgg atgtgggggc 660

tggggagtgg gccttgtcca gagatggggt tattgtgttc attgtaccag ggttttcaat 720

ctcccagaaa agcaattctt gccagcacat ttaaaccatt tcacaaatgc aggcgattct 780

ggccaaaagg atccgaatca gaggtccagt tcaaaacctc ccatgaaaaa agtaataggc 840

aaaaagcgtg ctcatgccga agatgatgag ttggatccaa tggacccgag ctcttactca 900

gatgctccac ggggaggctg ggttgttgga ctgaaaggag tacaaccgcg agccgctgat 960

acaactgctt cgggtcctct gtttcaacaa cgaccatatc catcacctgg agctgttctg 1020

aggagaaacg cagaggtggc atcatcacag aagaagaaac caacctctca gttcacagaa 1080

attacaaaga gaggcgatgg aagtgatggt cttggagacg cagattgagt gagagcttcc 1140

ctcatcttat aagttggtta ccgtgtcttc ttgtgctttg tctgttgctt tctccatata 1200

tgagatcaag gagagtagtt tgtttacttt ttagtgaata caaacaaacc atacttcttc 1260

atgtgaattt tttttaacac cagattcaaa gtatgtatac atataattac tcagtagcaa 1320

gcaacggtct tcaacaaaaa aaaaaaaa 1348

<210> 36
<211> 1612
<212> DNA
<213> Arabidopsis thaliana

<400> 36
atgggagaag agctgcaata tcagcagaat ggttcatctg tgacatctaa ttatggttat 60

ggaagttcgc ttgcttacga tcaatctcaa gacattgaat ctgccgcaaa cactgctctt 120

ctccgagaac aagaaataga aacgcagaag attatacaag gtcaaagaga agctgggact 180

tcagtagccg gagactctaa gcataacact gatattcttc gagatcgtgc tgaccccaat 240

gctctaaagg aacatttgct aaagttcact gcaaatcata gagcagaagc agcggcaaag 300

cgtggtggga gtgtgagtac atgcggagaa ggtaatgtgg atgttggaaa cggttatggt 360

ataccgggtg gagttgcata cgcgggtcac tccgagctca gtgggaagcc tgaacctacc 420

aatgcttcaa acaatttgcc tgaatatctc aaacaaaaat taaaggccag gggaattctc 480

agagatggtg caggggctgt tacatccaac ccagaagata catcagcggt tagttggaat 540

agacaggcaa ctttaccttt tcaggctaat gcaagcacct taccattagg atgggtagat 600

gcaaaagatc ctgccagtgg tgctacctat tactataacc agcacactgg aacatgccag 660

tgggaaaggc ctgttgagct ttcttatgcc acttcaagtg caccacctgt actttctaaa 720

gaagagtgga ttgaaacatt tgatgaagca tctggccata agtatttcta caatacaagg 780

acacatgtgt ctcagtggga acctccagct tctttacaga agcccgctgc cacaaactct 840

aacaacgctg ttacccaaag tacggctaat gggaaggggg agcatcctcc atctcagctg 900

ccaagatgca gcggatgtgg gggctgggga gtgggccttg tccagagatg gggttattgt 960

gttcattgta ccagggtttt caatctccca gaaaagcaat tcttgccagc acatttaaac 1020

catttcacaa atgcaggcga ttctggccaa aaggatccga atcagaggtc cagttcaaaa 1080

cctcccatga aaaaagtaat aggcaaaaag cgtgctcatg ccgaagatga tgagttggat 1140

ccaatggacc cgagctctta ctcagatgct ccacggggag gctgggttgt tggactgaaa 1200

ggagtacaac cgcgagccgc tgatacaact gcttcgggtc ctctgtttca acaacgacca 1260

tatccatcac ctggagctgt tctgaggaga aacgcagagg tggcatcatc acagaagaag 1320

aaaccaaact ctcagttcac agaaattaca aagagaggcg atggaagtga tggtcttgga 1380

gacgcagatt gagtgagagc ttccctcatc ttataagttg gttaccgtgt cttcttgtgc 1440

tttgtctgtt gctttctcca tatatgagat caaggagagt agtttgttta ctttttagtg 1500

aatacaaaca aaccatactt cttcatgtga attttttta acaccagatt caaagtatgt 1560

atacatataa ttactcagta gcaagcaacg gtcttcaaca aaaaaaaaaa aa 1612

<210> 37
<211> 3477
<212> DNA
<213> Arabidopsis thaliana

<400> 37
atgggagaag agctgcaata tcagcagaat ggttcatctg tgacatctaa ttatggttat 60

ggaagttcgc ttgcttacga tcaatctcaa gacattgaat ctgccgcaaa cactgctctt 120

ctccgagaac aagtacactt ttttttgtct tcttcttctt caatataagc ttctaaattc 180

cattcccgga ctttgccaat ggttaccatg ttgatttggt tttgtgtgtt ttcaggaaat 240

agaaacgcag aagattatac aaggtcaaag gttctgtctt tctcttggtt tgttctttat 300

atccatctgt tatatatgct taatgcttgt attgatcatc ttgctgtttc tcattcttgg 360

gcaatgagaa atttcgtgtg tttaggtgct ttgtctaaca atttgtgttt aattggaaac 420

gttctcctat tttcagagaa gctgggactt cagtagccgg agactctaag cataacactg 480

atattcttcg agatcgtgct gaccccaatg ctctaaaggt tctgattcac tatttttttt 540

gtttatcatc aagcggctat tggcttaata agtcctataa tcacatgaag ttttgataca 600

tattatcttt atgcaggaac atttgctaaa gttcactgca aatcatagag cagaagcagc 660

ggcaaagcgt ggtgggagtg tgagtacatg cggagaaggt tagtcataat agcttacgag 720

ctcagtcatt taggaaaatt caggaaactg atccaattat cgactttatt ccgttctagg 780

taatgtggat gttggaaacg gttatggtat accgggtgga gttgcatacg cgggtcactc 840

cgagctcagt gggaagcctg aacctaccaa tgcttcaaac aatttgcctg aatatctcaa 900

acaaaaatta aaggccaggg gaattctcag agatggtgca ggggctgtta catccaaccc 960

agaagatgta attattcatc cctcagactg taatccaggg tgacatatgt ctgcaatgac 1020

ttgtccatat ttcagcccct ttgtttcagt ttatcttatt aatatttgtt aacctttga 1080

cgtttttgtt tgcttgtgtt tcttctcttt tccagacatc agcggttagt tggaatagac 1140

aggcaacttt acctttcag gctaatgcaa gcaccttacc attaggatgg gtatgaatga 1200

gatacactta aaagtctttt cttcttcatt tttgtctttg ctttactatc tggtgcagca 1260

atattttctg ctgtgaaagt gcggtaatcg tttgagattc tccaaatgca tgtgataatg 1320

tggtcataca caaatagtct ttagatctca atccttaagg atgaccactt atggttctga 1380

tgatgtgttt atagttgtta taccacttaa atttgttctg tttcattcag gtagatgcaa 1440

aagatcctgc cagtggtgct acctattact ataaccagca cactggaaca tgccagtggg 1500

aaaggcctgt tgagctttct tatgccactt caagtgcacc acctgtactt tctaaagaag 1560

agtggattga aacatttgat gaagcatctg gtatgatttt cacacattta gtcatcataa 1620

tgactagctc gttgctacaa tattcatatc aaagattata aaagtatgat atctttctcc 1680

ttttcttatt gtggtattat gctgttagca aaatgaattt aatagcttat gatctttttc 1740

atcttggaaa gaaacttaca tcatcacatg ttagaaataa gaatgaggaa atccttgtgt 1800

tttctgctgg aagtttttgc ctcttattga ccagcttctt ctgatttctg ttgtactgta 1860

cacgcaatgc ttgcttaatc tttggatctg ttggttgcat ctttgaactc tgaacctttt 1920

gattcctatg cacaggccat aagtatttct acaatacaag gacacatgtg tctcagtggg 1980

aacctccagc ttctttacag aagcccgctg ccacaaactc taacaacgct gttacccaaa 2040

gtacggctaa tgggaagggg gagcatcctc catctcagct gccaagatgc agcggatgtg 2100

ggggctgggg agtgggcctt gtccagagat ggggttattg tgttcattgt accaggtagg 2160

aacccttcca tttcttgtta tgctttttga gttagattgc tgcatagaga ctgcttcggt 2220

tgaagactgt taataatgta tatacctcta gtcctctacc atgactccct ttttattact 2280

ggctcagatt tgtgttatag gaagcattta ttagtgtgca tcaagacctt aaatacactt 2340

tgtttacttg catattttga ggcttaagcc cctctgcctt tggttaacgg gactcatatt 2400

atatgggatt cagggttttc aatctcccag aaaagcaatt cttgccagca catttaaacc 2460

atttcacaaa tgcaggcgat tctggccaaa aggatccgaa tcagaggtac gcatctctca 2520

aaatcttaga tttagaaaaa atattcttct atgcccgtca tctttctaaa atacatagtg 2580

ttccataaga gagctattgg tggactctgg tcatcttctt aacatctgtt accagatatt 2640

tagacaaccc atttggatac aaatgttgcc cctcttttcg gagggacata atttctgtta 2700

aaaagcattc catttatcag cagtaaaatg aagatggaac acattacatg tagtagtgaa 2760

ccaggaggtt cccattgact atgtttacat tgtacttact cttgtggtat attctttaca 2820

catgatggaa aaatgaaaat ggttgaaaca ggtccagttc aaaacctccc atgaaaaaag 2880

taataggcaa aaagcgtgct catgccgaag atgatgagtt ggatccaatg gacccgagct 2940

cttactcaga tgctccacgg ggaggctggt aatgtccata acacccattt taaccacttc 3000

atgaaacggc atgacctata atactgttct ctatctctct ctgctatgtc atctgaattg 3060

aggagaacca agttcttttt tggttaacaa ccaagctaat tatgctcata ggaatcataa 3120

tatcaggttg atctgattgc ttctcatctt gtagggttgt tggactgaaa ggagtacaac 3180

cgcgagccgc tgatacaact gcttcggtaa tggcatttac cttttcagac tgggatcttc 3240

tttaaattcc aaataaccga cctcactaaa ccagatgaaa aaattgaggt tctaactaaa 3300

cgagatcaaa tcgtgtttca gggtcctctg tttcaacaac gaccatatcc atcacctgga 3360

gctgttctga ggagaaacgc agaggtggca tcatcacaga agaagaaacc aaactctcag 3420

ttcacagaaa ttacaaagag aggcgatgga agtgatggtc ttggagacgc agattga 3477

<210> 38
<211> 1500
<212> DNA
<213> Arabidopsis thaliana

<400> 38
aaattggatc tgattgttgg ttcagcgcca acatacatct cttctgcatc tcaaatggct 60

gatggggcac ggatgttgat tcctcaacgg ggccttaacg atatcgtggt tcctgtatac 120

gatgatgatc cagctagtgt tgtatcatat gctatcaatt caaaggaata taaggagtgg 180

atcgttaaca aaggtcttgc ttctagtagt agtagcagca atttgaataa cagagagtct 240

gaaccttccg ccttctcgac ttggcgttct ctgtcgatgg atgttgatta cattcaacac 300

gcggtgtatg gatcttctca agacgataga aagtctccgc atttgaccat ttctttcagc 360

gaccgcgctt cttcttcctc tactgccact gaaggtaaag tgaagttctc tgtgacgtgt 420

tattttgcaa cgcagtttga cactctgaga aagacttgct gtccgagtga agtggacttt 480

gtgagatcct tgagccgttg tcagagatgg tctgcacagg gagggaaaag caatgtttac 540

tttgccaagt cattggacga gaggttcatc ataaaacaag tcgtcaaaac cgagctggat 600

tctttcgagg attttgcacc tgagtacttc aaatatttga aggaatcact cagttccggg 660

agccctactt gtcttgctaa gattctcggt atctaccagg tctcaattaa acatccgaaa 720

ggtggaaaag agacaaagat ggatttgatg gtgatggaga atctcttcta caacagaaga 780

atatctagga tctatgatct caagggatct gcacggtcaa ggtacaatcc aaacacatcc 840

ggagcagaca aagttttgtt agacatgaat ctgctagaga cactgcgcac agaaccgata 900

ttcctaggaa gcaaggccaa gagaagcttg gaaagagcga tatggaacga caccaatttc 960

ttggcttctg tggatgtaat ggactactca ttgctggttg gattcgacga agagcggaaa 1020

gaactagttc tcgggatcat cgacttcatg agacagtaca catgggacaa gcacctagag 1080

acttgggtta aagcttcagg cattttgggt ggacccaaga atgcttctcc aactatagtc 1140

tcaccaaaac agtacaagag gaggtttaga aaggccatga ccacttattt tctcactgtt 1200

cctgagccat ggacctcttg agtgaaagaa aagaaaccat ttttcatcat tcttttgttc 1260

actgattcca atttgttctt ctcaaataaa aagggaaatt ggttaacaga cacatataat 1320

ataattcagg ttcgaacctg actgattgat ttatcctttt ttttttttt ttttggtgt 1380

gttggtttct ctgagttctg gagaattgta tattgtaatt atcggccaaa ctaaagcccc 1440

ccaaatttaa agaaaaagaa tactaatgat ttgcctttta aaaaaaaaaa aaaaaaaaaa 1500

<210> 39
<211> 5226
<212> DNA
<213> Arabidopsis thaliana

<400> 39
atgggaatac ctgatggttc actgctagat ctaattgata aggttaggtc ttggataact 60

tcggattcaa gtgattcttt gttcttgttg tcttcatcta aacaagactt tggaattatg 120

cctatcgttt ccaaaatgtg tcatgactgt ggaacaaagg tggaacaagg atattgttgt 180

cttagttgtg gttcttgttg gtgcaagagt tgttctgata cagaagagag caagatgaag 240

ttgtgtagag aatgtgatgc tgaggttcgt gagttaaggg ttaagagcta tgataaggtt 300

catcctcggg atagccctga tccgccttct tctcttgcta ccgaaagcga gagtcttgcc 360

tcgagtcttg aaatccgtga ttgtaggaac atggcttcta ttcgttgcta tcctagcagg 420

ggagaggaag aagaggctag atattgtggt aaacagttgt taagcccttc gagtgataac 480

tatcaggaca gttcagatat agagtcgggt agtgttagtg ctagacatga gctttttagt 540

tgtaaatctt ctgctgggtc cagtcctcac gacagtccat taaggaacaa ttttagtcct 600

cttgggcgct ttgtacagca cgcgaaagac ctaagaagcc ctactgtttg ttcttttgat 660

aatcatcaag aacaactctt ggctgataat ttagttaagc caggtcaggg ggtattggaa 720

caggaagatc acgaggagga agaggataag ttgcaacagc cgttggattt tgagaacaac 780

ggccgcatct ggtatcctcc accaccggaa gatgagaatg atgatgctga gagtaattac 840

tttcattatg atgacgaaga tgacgatatt ggggactctg ctaccgagtt ttcattgagt 900

agtagctttt ctagccacat tcctaccaaa gagaaattgg gagaaaacag caatgaacct 960

cttcgaaccg tggtgcatga ccacttccga gctcttgttg cagagctatt acgtggggag 1020

gagcttagtc catctgatga tggcagtgcc ggggagtggc tggatattgt cactgcttta 1080

gcgtggcaag ctgcaaattt tgtaaagccg gatactcgtg caggcggcag tatggatccc 1140

gggaattatg taaaaataaa atgtgtagca tctggaaatc aaaatgagag cattctcatc 1200

aggggaatag tgtgtagcaa gaacattacg cacaagcgaa tgatatcgca gtacaaaaat 1260

ccaagggtga tgcttttagc tggttctctt gagtatcaga gagttgcagg ccaattagca 1320

tcctttaaca cgttgcttca acaggaaaat gaacatatga aggcgataat agcaaaaata 1380

gagtcccttc gtcctaatgt tctgctggta gagaaaagtg catcttcata tgcccagcaa 1440

taccttctag agaaggaaat ttcattggtg ctcaatgtga aaaggtcatt gctggaccgt 1500

atagcccgtt gcactggtgc tgttctttgc ccctcactag atagtattag tactgctcga 1560

ttgggtcact gtgagctgtt ccggactgag agagtgttgg agcaacatga agctggcaat 1620

cagtcaaaca gaaaaccatc aaggacattg atgtactttg aagggtgtcc taggcgccta 1680

ggttgcacgg ttgtgcttcg gggaagttgt cgtgaggagc taaagaaggt taaacatgtt 1740

attcagtacg ctgtatttgc agcttatcac ttatcgttgg agacttcatt ccttgcggat 1800

gaaggtgctt ctctgcctaa aattagactg aagcaaccag ggatggtgcg gactgcatca 1860

caaaggagga taattgatga aggcatttcc ttaataactc aatctcctac agaaacagat 1920

agtcaagctt tgcttgaaac agctgcacat gaagatgaac atacggctcc tatgcctgag 1980

catgaagttt gtgaatcatt gtgtgaggat tttgatccta ctcagatatt cccaccgtct 2040

tctgaagttg agactgaaca atctgatacc ctgaatggag attttgccaa taatttggta 2100

acaagatcat attcttcgaa tcagttaaat gatttgcatg aacccacttt gtgcttaagt 2160

agtgaaattc ctgaaacgcc gactcaacag ccgagtggtg aagaagataa cggcagaggt 2220

gaagaagaga accaattagt gaacccccaa gatttgccac aacatgagag cttctatgaa 2280

gatgatgtct ctagtgagta tttctctgct gcggactctc atcagagcat attggtctca 2340

ttttcaagcc gttgtgtttt gaaagaatca gtatgtgaac gttcacgact cttgcggatt 2400

aagttctatg gatcctttga taaacctctc ggaagatatt tgaaagatga tcttttcgat 2460

aagacttcta gttgtagatc ctgtaaggag ctggttgatg cgcatgtact ctgctattct 2520

catcagaatg gaaaccttac tatcaatgtt agacgtctcc cttccatgaa gcttcctggt 2580

gaacaagatg ggaagatatg gatgtggcat cgttgcttaa gatgtgcaca tgtagatgga 2640

gtcccacctg ctactcgtag agtagttatg tctgatgcag catggggatt gtcttttgga 2700

aagtttctgg aacttagctt ttcgaatcat gcaactgcga atcgtgttgc atcttgcggt 2760

cattcgcttc aaagagactg ccttcgattc tatggatttg gtaacatggt ggcattcttt 2820

agatattctc ccatcaatat acttactgta ctcctcccac cgtcgatgct tgagttcaat 2880

agccatcctc aacaggagtg gatacgaaca gaggcagctg agcttgttgg taaaatgagg 2940

actatgtata cggagatatc tgatatgctc aaccgcatgg aagaaaaaag cagtttgctt 3000

gaacctgaac aatctgaagc atgtgatcta catagtcgca tcatcggatt aatagatcaa 3060

cttgtgaaag aaaaagatga atacgatgat gcactgcaac ctatttttga ggagaatctg 3120

caaattcagg gaagtttgga tatcctagag ctgaatcgat tgagacgagc actcatgatt 3180

ggtgctcatg cttgggatca tcagctttac ttgttaaact ctcaactcaa gaaagcatca 3240

gtctttaaaa ctggtgatga caatgctcct cggaatccag agatgcatga tcctccaaag 3300

attgaccgta gaatgcagga aggttcagat gagagggatg aacaatccca tactgattca 3360

gaagctaatg gtgataacaa agatccggaa aacattccct cccctggcac ttctctatct 3420

gagaggatag attcggcgtg gttgggctcg tttcaaaatc tagagaaagc cgagactatt 3480

gctgagacgg aaggcttttc agctgttaat tcttctctcc ggaggcttgc caggccaatt 3540

cgagtccagt cttttgattc agcgatacgg tttcaagaac gtatccaaaa aggtttgcca 3600

ccatcttctt tgtatctctc gaccctcaga tctttccacg cttctggcga gtataggaat 3660

atggtgaggg accctgtttc aaatgtgatg aggacttact cacagatgtt gccactggaa 3720

gtacagaaat tggatctgat tgttggttca gcgccaacat acatctcttc tgcatctcaa 3780

atggctgatg gggcacggat gttgattcct caacggggcc ttaacgatat cgtggttcct 3840

gtatacgatg atgatccagc tagtgttgta tcatatgcta tcaattcaaa ggaatataag 3900

gagtggatcg ttaacaaagg tcttgcttct agtagtagta gcagcaattt gaataacaga 3960

gagtctgaac cttccgcctt ctcgacttgg cgttctctgt cgatggatgt tgattacatt 4020

caacacgcgg tgtatggatc ttctcaagac gatagaaagt ctccgcattt gaccatttct 4080

ttcagcgacc gcgcttcttc ttcctctact gccactgaag gtaaagtgaa gttctctgtg 4140

acgtgttatt ttgcaacgca gtttgacact ctgagaaaga cttgctgtcc gagtgaagtg 4200

gactttgtga gatccttgag ccgttgtcag agatggtctg cacagggagg gaaaagcaat 4260

gtttactttg ccaagtcatt ggacgagagg ttcatcataa aacaagtcgt caaaaccgag 4320

ctggattctt tcgaggattt tgcacctgag tacttcaaat atttgaagga atcactcagt 4380

tccgggagcc ctacttgtct tgctaagatt ctcggtatct accaggtctc aattaaacat 4440

ccgaaaggtg gaaaagagac aaagatggat ttgatggtga tggagaatct cttctacaac 4500

agaagaatat ctaggatcta tgatctcaag ggatctgcac ggtcaaggta caatccaaac 4560

acatccggag cagacaaagt tttgttagac atgaatctgc tagagacact gcgcacagaa 4620

ccgatattcc taggaagcaa ggccaagaga agcttggaaa gagcgatatg gaacgacacc 4680

aatttcttgg cttctgtgga tgtaatggac tactcattgc tggttggatt cgacgaagag 4740

cggaaagaac tagttctcgg gatcatcgac ttcatgagac agtacacatg ggacaagcac 4800

ctagagactt gggttaaagc ttcaggcatt ttgggtggac ccaagaatgc ttctccaact 4860

atagtctcac caaaacagta caagaggagg tttagaaagg ccatgaccac ttattttctc 4920

actgttcctg agccatggac ctcttgagtg aaagaaaaga aaccattttt catcattctt 4980

ttgttcactg attccaattt gttcttctca ataaaaagg gaaattggtt aacagacaca 5040

tataatataa ttcaggttcg aacctgactg attgatttat ccttttttt ttttttttt 5100

tggtgtgttg gtttctctga gttctggaga attgtatatt gtaattatcg gccaaactaa 5160

agccccccaa atttaaagaa aaagaatact aatgatttgc cttttaaaaa aaaaaaaaaa 5220

aaaaaa 5226

<210> 40
<211> 5874
<212> DNA
<213> Arabidopsis thaliana

<400> 40
atgggaatac ctgatggttc actgctagat ctaattgata aggttaggtc ttggataact 60

tcggattcaa gtgattcttt gttcttgttg tcttcatcta aacaagactt tggaattatg 120

cctatcgttt ccaaaatgtg tcatgactgt ggaacaaagg tggaacaagg atattgttgt 180

cttagttgtg gttcttgttg gtgcaagagt tgttctgata cagaagagag caagatgaag 240

ttgtgtagag aatgtgatgc tgaggttcgt gagttaaggg ttaagagcta tgataaggtt 300

catcctcggg atagccctga tccgccttct tctcttgcta ccgaaagcga gagtcttgcc 360

tcgagtcttg aaatccgtga ttgtaggaac atggcttcta ttcgttgcta tcctagcagg 420

taaagtttcg atttttacat tcaattatgt gtgaaaaaga ttcaattttt atgttcaatt 480

atgagtcttt aaacatgttg gagtaagtga atgatttgtg tttgcagggg agaggaagaa 540

gaggctagat attgtggtaa acagttgtta agcccttcga gtgataacta tcaggacagt 600

tcagatatag agtcgggtag tgttagtgct agacatgagc tttttagttg taaatcttct 660

gctgggtcca gtcctcacga cagtccatta aggaacaatt ttagtcctct tgggcgcttt 720

gtacagcacg cgaaagacct aagaagccct actgtttgtt cttttgataa tcatcaagaa 780

caactcttgg ctgataattt agttaagcca ggtcaggggg tattggaaca ggaagatcac 840

gaggaggaag aggataagtt gcaacagccg ttggattttg agaacaacgg ccgcatctgg 900

tatcctccac caccggaaga tgagaatgat gatgctgaga gtaattactt tcattatgat 960

gacgaagatg acgatattgg ggactctgct accgagtttt cattgagtag tagcttttct 1020

agccacattc ctaccaaaga gaaattggga gaaaacagca atgaacctct tcgaaccgtg 1080

gtgcatgacc acttccgagc tcttgttgca gagctattac gtggggagga gcttagtcca 1140

tctgatgatg gcagtgccgg ggagtggctg gatattgtca ctgctttagc gtggcaagct 1200

gcaaattttg taaagccgga tactcgtgca ggcggcagta tggatcccgg gaattatgta 1260

aaaataaaat gtgtagcatc tggaaatcaa aatgagaggt ataactagtt gttgaaacct 1320

tgtttaaatc gtcctctgac tgatgattac taatttgggt gatagttact aaatgaatac 1380

cttgtttctg cagcattctc atcaggggaa tagtgtgtag caagaacatt acgcacaagc 1440

gaatgatatc gcagtacaaa aatccaaggg tgatgctttt agctggttct cttgagtatc 1500

agagagttgc aggccaatta gcatccttta acacgttgct tcaacaggtt ggtgtgttac 1560

tgtttcttta tgatggctgc ataaaatctc tgattaacaa actgaacaac tgaattatgt 1620

gtgttggaca ggaaaatgaa catatgaagg cgataatagc aaaaatagag tcccttcgtc 1680

ctaatgttct gctggtagag aaaagtgcat cttcatatgc ccagcaatac cttctagaga 1740

aggaaatttc attggtgctc aatgtgaaaa ggtcattgct ggaccgtata gcccgttgca 1800

ctggtgctgt tctttgcccc tcactagata gtattagtac tgctcgattg ggtcactgtg 1860

agctgttccg gactgagaga gtgttggagc aacatgaagc tggcaatcag tcaaacagaa 1920

aaccatcaag gacattgatg tactttgaag ggtgtcctag gcgcctaggt tgcacggtat 1980

gctcctttat tggtttcaca gttcttgttt tgcatagctt gaacctttgt agctaagaag 2040

tctgctatct taaatttatt aacttcctta ctgacttttc tgaaacttgt gcaggttgtg 2100

cttcggggaa gttgtcgtga ggagctaaag aaggttaaac atgttattca gtacgctgta 2160

tttgcagctt atcacttatc gttggagact tcattccttg cggatgaagg tgcttctctg 2220

cctaaaatta gactgaagca accagggatg gtgcggactg catcacaaag gaggataatt 2280

gatgaaggca tttccttaat aactcaatct cctacagaaa cagatagtca agctttgctt 2340

gaaacagctg cacatgaaga tgaacatacg gctcctatgc ctgagcatga agtttgtgaa 2400

tcattgtgtg aggattttga tcctactcag atattcccac cgtcttctga agttgagact 2460

gaacaatctg ataccctgaa tggagatttt gccaataatt tggtaacaag atcatattct 2520

tcgaatcagt taaatgattt gcatgaaccc actttgtgct taagtagtga aattcctgaa 2580

acgccgactc aacagccgag tggtgaagaa gataacggca gaggtgaaga agagaaccaa 2640

ttagtgaacc cccaagattt gccacaacat gagagcttct atgaagatga tgtctctagt 2700

gagtatttct ctgctgcgga ctctcatcag agcatattgg tctcattttc aagccgttgt 2760

gttttgaaag aatcagtatg tgaacgttca cgactcttgc ggattaagtt ctatggatcc 2820

tttgataaac ctctcggaag atatttgaaa gatgatcttt tcgataaggt agctctttca 2880

tctataaagt tagttaatcc actttttcat ctttcttaaa ttttgaagtc taatgccttt 2940

ctgttccatt tctggacaga cttctagttg tagatcctgt aaggagctgg ttgatgcgca 3000

tgtactctgc tattctcatc agaatggaaa ccttactatc aatgttagac gtctcccttc 3060

catgaagctt cctggtgaac aagatgggaa gatatggatg tggcatcgtt gcttaagatg 3120

tgcacatgta gatggagtcc cacctgctac tcgtagagta gttatgtctg atgcagcatg 3180

gggattgtct tttggaaagt ttctggaact tagcttttcg aatcatgcaa ctgcgaatcg 3240

tgttgcatct tgcggtcatt cgcttcaaag agactgcctt cgattctatg ggtaaacgca 3300

ggctgtataa tatcattctc ttacagttat ctcaatgctg ttacactaat aatattatga 3360

tttgttttgt agatttggta acatggtggc attctttaga tattctccca tcaatatact 3420

tactgtactc ctcccaccgt cgatgcttga gttcaatagc catcctcaac aggagtggat 3480

acgaacagag gcagctgagg tagattttaa gaaaaactta aggctctttt gcttcaggtt 3540

ttggatatgt aaatgctaaa actcaaatac ttctctgttt gcattttaca gcttgttggt 3600

aaaatgagga ctatgtatac ggagatatct gatatgctca accgcatgga agaaaaaagc 3660

agtttgcttg aacctgaaca atctgaagca tgtgatctac atagtcgcat catcggatta 3720

atagatcaac ttgtgaaaga aaaagatgaa tacgatgtaa gcttcttctg tttctgaaat 3780

cttgatgatg cttgttgaat atgtatttga tggaatattc ctcatcgtga acaggatgca 3840

ctgcaaccta tttttgagga gaatctgcaa attcagggaa gtttggatat cctagagctg 3900

aatcgattga gacgagcact catgattggt gctcatgctt gggatcatca gctttacttg 3960

ttaaactctc aactcaagaa agcatcagtc tttaaaactg gtgatgacaa tgctcctcgg 4020

aatccagaga tgcatgatcc tccaaagatt gaccgtagaa tgcaggaagg ttcagatgag 4080

agggatgaac aatcccatac tgattcagaa gctaatggtg ataacaaaga tccggaaaac 4140

attccctccc ctggcacttc tctatctgag aggatagatt cggcgtggtt gggctcgttt 4200

caaaatctag agaaagccga gactattgct gagacggaag gcttttcagc tgttaattct 4260

tctctccgga ggcttgccag gccaattcga gtccagtctt ttgattcagc gatacggttt 4320

caagaacgta tccaaaaagg tttgccacca tcttctttgt atctctcgac cctcagatct 4380

ttccacgctt ctggcgagta taggaatatg gtgagggacc ctgtttcaaa tgtgatgagg 4440

acttactcac agatgttgcc actggaagta cagaaattgg atctgattgt tggttcagcg 4500

ccaacataca tctcttctgc atctcaaatg gctgatgggg cacggatgtt gattcctcaa 4560

cggggcctta acgatatcgt ggttcctgta tacgatgatg atccagctag tgttgtatca 4620

tatgctatca attcaaagga atataaggag tggatcgtta acaaaggtct tgcttctagt 4680

agtagtagca gcaatttgaa taacagagag tctgaacctt ccgccttctc gacttggcgt 4740

tctctgtcga tggatgttga ttacattcaa cacgcggtgt atggatcttc tcaagacgat 4800

agaaagtctc cgcatttgac catttctttc agcgaccgcg cttcttcttc ctctactgcc 4860

actgaaggta aagtgaagtt ctctgtgacg tgttattttg caacgcagtt tgacactctg 4920

agaaagactt gctgtccgag tgaagtggac tttgtgagat ccttgagccg ttgtcagaga 4980

tggtctgcac agggagggaa aagcaatgtt tactttgcca agtcattgga cgagaggttc 5040

atcataaaac aagtcgtcaa aaccgagctg gattctttcg aggattttgc acctgagtac 5100

ttcaaatatt tgaaggaatc actcagttcc gggagcccta cttgtcttgc taagattctc 5160

ggtatctacc aggttagtga caatatagga gggaatgtga acttgtagga ctctgaaatg 5220

aaataagtga gcatgtgaat gattgatttg gaatgatatt tgttgtaggt ctcaattaaa 5280

catccgaaag gtggaaaaga gacaaagatg gatttgatgg tgatggagaa tctcttctac 5340

aacagaagaa tatctaggat ctatgatctc aagggatctg cacggtcaag gtacaatcca 5400

aacacatccg gagcagacaa agttttgtta gacatgaatc tgctagagac actgcgcaca 5460

gaaccgatat tcctaggaag caaggccaag agaagcttgg aaagagcgat atggaacgac 5520

accaatttct tggctgtaag tgttttcatt acccccaatt catattttgt tcttaagtgt 5580

tttgattaat ctgtttcccc attatcctct tgtggttagt ctgtggatgt aatggactac 5640

tcattgctgg ttggattcga cgaagagcgg aaagaactag ttctcgggat catcgacttc 5700

atgagacagt acacatggga caagcaccta gagacttggg ttaaagcttc aggcattttg 5760

ggtggaccca agaatgcttc tccaactata gtctcaccaa aacagtacaa gaggaggttt 5820

agaaaggcca tgaccactta ttttctcact gttcctgagc catggacctc ttga 5874

<210> 41
<211> 643
<212> DNA
<213> Arabidopsis thaliana

<400> 41
ggcacgagag acaaagccaa ggaagtttca gaactgaaag cagagaggca gaagaagaaa 60

cttcagatag atgagctaga gagaatcgtg aggctgaagc aagcggaggc agacatgttc 120

cagcttaaag caaatgaagc aaagcgggag gctgataggt tgcagaggat tgtactagcg 180

aaaatggata agtctgagga ggaatacgca agtaactatc tgaaacagag gctgagtgag 240

gccgaggcag agaagcagta tctgtttgag aagattaagc tgcaggaaaa ctcaagggtg 300

gcatcacaaa gcagtggtgg gggaggagac ccgtcacagg tgatgatgta ctcaaagatc 360

cgcgatctgc ttcaaggata caatctctct cccaaggttg atcctcaatt aaacgagcgt 420

aatcctttca gatccaatcc ctaatgcttc ttcctccctt atatatatat gtagttttca 480

tatcctaaat ttcatcactg tgttttgtaa aacgagtgtt tgttgccggt aggatctatc 540

tatattactg atgttttgta atttgtgtgt tctcactttg ttgatatgac gttgtttatc 600

tatgttatgg tttcttggtt ggttgaaaaa aaaaaaaaaa aaa 643

<210> 42
<211> 1924
<212> DNA
<213> Arabidopsis thaliana

<400> 42
atgggaacat catctggttc gaatcatcct caccaaatgt taccaccacg tcaacaacaa 60

cgaagtggag gaggacttga aacagctctt tcactagtct cttcagatca agagccgcgt 120

cgcgagtctc cagcggagag tgcgagttct caagaaacat ggccactagg agacactgtt 180

gctgggaaga agtcaatgag tcaaaagaca gagcctgatt ctatggaaca aactgtgaat 240

gtgatgcacc atgtttcgaa tgcggacaag gtatcggtac gggatatcgc tagggaaaga 300

gttgagttag ttgcagagag aatgcatcga ttaccggatg agtttcttga cgagttaaag 360

aacggtctta agtctatcct tgaagggaat gtagcgcaga gtgttgatga gttcatgttc 420

ttgcagaagg ttgttcagag tagaactgat ttgagctctg taacacttgt tagagctcat 480

cgggtgcagc ttgagatact tgtagcgata aataccggga ttcaggcgtt tttgcatccg 540

aatattagtc tgtctcagcc gtctcttatc gagatttttg tatacaagag atgtagaaac 600

atagcttgcc aaaatcaact acctgctgat gattgctact gtgatatatg tactaatagg 660

aaagggtttt gtaacctttg tatgtgtacg atctgtaaca agtttgattt ttcggttaat 720

acttgccgct ggatcggatg tgacttgtgt tcccactgga ctcatacgga ttgtgctatt 780

cgggatggac agatcaccac tggatcctct gctaagaata atacttcagg tccaggagag 840

atagtgttta agtgtcgggc ttgtaaccga acgtctgagc tgttgggatg ggttaaagat 900

gtgtttcagc actgtgcacc gaactgggat cgagaatctt tgatgaagga acttgatttt 960

gttagtagga ttttccgcgg aagcgaagat caacgaggta ggaaactgtt ttggaagtgt 1020

gaggagctta ttgacaagat taaaggtggt ttggctgaag caactgctgc caaattgata 1080

ttaatgtttt tccaagagat cgaatcagac tctgcgaaga gctttgaaaa tggagaaggt 1140

ggacgtttga tggcacctca ggatgcatgc aaccgaattg ctgaagttgt acaagagact 1200

ttgaggaaaa tggagatagt ggctgaggag aagatgagaa tgttcaagaa ggcacgcatg 1260

gctcttgaga catgcgatag agagctcgaa gacaaagcca aggaagtttc agaactgaaa 1320

gcagagaggc agaagaagaa acttcagata gatgagctag agagaatcgt gaggctgaag 1380

caagcggagg cagacatgtt ccagcttaaa gcaaatgaag caaagcggga ggctgatagg 1440

ttgcagagga ttgtactagc gaaaatggat aagtctgagg aggaatacgc aagtaactat 1500

ctgaaacaga ggctgagtga ggccgaggca gagaagcagt atctgtttga gaagattaag 1560

ctgcaggaaa actcaagggt ggcatcacaa agcagtggtg ggggaggaga cccgtcacag 1620

gtgatgatgt actcaaagat ccgcgatctg cttcaaggat acaatctctc tcccaaggtt 1680

gatcctcaat taaacgagcg taatcctttc agatccaatc cctaatgctt cttcctccct 1740

tatatatata tgtagttttc atatcctaaa tttcatcact gtgttttgta aaacgagtgt 1800

ttgttgccgg taggatctat ctatattact gatgttttgt aatttgtgtg ttctcacttt 1860

gttgatatga cgttgtttat ctatgttatg gtttcttggt tggttgaaaa aaaaaaaaaa 1920

aaaa 1924

<210> 43
<211> 1802
<212> DNA
<213> Arabidopsis thaliana

<400> 43
atgggaacat catctggttc gaatcatcct caccaaatgt taccaccacg tcaacaacaa 60

cgaagtggag gaggacttga aacagctctt tcactagtct cttcagatca agagccgcgt 120

cgcgagtctc cagcggagag tgcgagttct caagaaacat ggccactagg agacactgtt 180

gctgggaaga agtcaatgag tcaaaagaca gagcctgatt ctatggaaca aactgtgaat 240

gtgatgcacc atgtttcgaa tgcggacaag gtatcggtac gggatatcgc tagggaaaga 300

gttgagttag ttgcagagag aatgcatcga ttaccggatg agtttcttga cgagttaaag 360

aacggtctta agtctatcct tgaagggaat gtagcgcaga gtgttgatga gttcatgttc 420

ttgcagaagg ttgttcagag tagaactgat ttgagctctg taacacttgt tagagctcat 480

cgggtgcagc ttgagatact tgtagcgata aataccggga ttcaggcgtt tttgcatccg 540

aatattagtc tgtctcagcc gtctcttatc gagatttttg tatacaagag atgtagaaac 600

atagcttgcc aaaatcaact acctgctgat gattgctact gtgatatatg tactaatagg 660

aaagggtttt gtaacctttg tatgtgtacg atctgtaaca agtttgattt ttcggttaat 720

acttgccgct ggatcggatg tgacttgtgt tcccactgga ctcatacgga ttgtgctatt 780

cgggatggac agatcaccac tggatcctct gctaagaata atacttcagg tccaggagag 840

atagtgttta agtgtcgggc ttgtaaccga acgtctgagc tgttgggatg ggttaaagat 900

gtgtttcagc actgtgcacc gaactgggat cgagaatctt tgatgaagga acttgatttt 960

gttagtagga ttttccgcgg aagcgaagat caacgaggta ggaaactgtt ttggaagtgt 1020

gaggagctta ttgacaagat taaaggtggt ttggctgaag caactgctgc caaattgata 1080

ttaatgtttt tccaaggtaa cttctttttt atgaaccaac attactattt gtctgcagag 1140

tcttaaccct tttttgttac tgttttacca aagagatcga atcagactct gcgaagagct 1200

ttgaaaatgg agaaggtgga cgtttgatgg cacctcagga tgcatgcaac cgaattgctg 1260

aagttgtaca agagactttg aggaaaatgg agatagtggc tgaggagaag atgagaatgt 1320

tcaagaaggc acgcatggct cttgagacat gcgatagaga gctcgaagac aaagccaagg 1380

aagtttcaga actgaaagca gagaggcaga agaagaaact tcagatagat gagctagaga 1440

gaatcgtgag gctgaagcaa gcggaggcag acatgttcca gcttaaagca aatgaagcaa 1500

agcgggaggc tgataggttg cagaggattg tactagcgaa aatggataag tctgaggagg 1560

aatacgcaag taactatctg aaacagaggc tgagtgaggc cgaggcagag aagcagtatc 1620

tgtttgagaa gattaagctg caggaaaact caagggtggc atcacaaagc agtggtgggg 1680

gaggagaccc gtcacaggtg atgatgtact caaagatccg cgatctgctt caaggataca 1740

atctctctcc caaggttgat cctcaattaa acgagcgtaa tcctttcaga tccaatccct 1800

aa 1802

<210> 44
<211> 589
<212> DNA
<213> Arabidopsis thaliana

<400> 44
atagcaaaaa gcttaagtca ggaaaaccta gtggagttat ctgatgaaaa tgatgatcta 60

tgtcctgtgg agtgtgtcac tgagttcaag acagatgatg aattgcttag cgttcttgaa 120

aagtcgaaag aaactaattc tttggttgtg gttgattttt atcgcactgc atgtgggagt 180

tgtaaataca tagagcaggg cttctcaaaa ctgtgcaagc aatctggcca agaagctcct 240

gttatcttcc ttaagcataa tgtggtagat gaatatgatg aacaatctga agtcgcagaa 300

aggctccgta tcaaggcggt tcctctcttc cacttctaca aaaacggagt tctcttagaa 360

tcatttgcaa ctagagacaa ggagaggatc gacgcagcta ttctcaaata tacatcctcg 420

gaatcttgaa gaacattcga caaaaccgca ctctttgggt taccttccc aaatcattga 480

atatgatgta acttatattg aacaacacag catcatcctt cttcttgtaa gatgtcatac 540

tgaataagac agatgcaaca ttttgatcta aaaaaaaaaa aaaaaaaaa 589

<210> 45
<211> 968
<212> DNA
<213> Arabidopsis thaliana

<400> 45
atggctgaca aagtcgatac caagttcgtc tgggttccga tttccgtccc aagtaggacg 60

gttcaagcgc atctacagaa aaacggtttg tactttcaaa agaacgcaac ttttatttcg 120

aaagttcgat cctttgggtc catagtattg acgggccact caaccaacta cgagtgctat 180

ctttcacccc tgagttcact gaaagtgttc ctgcttggtt tgctgatgag ctattgcttg 240

tggagactca ctagtagggg aggcataaac actgttgaaa accatagatg ggtctggcac 300

aacaacggag tcaggttgtc gtttccaaga gccgagtctt ctataaacat cactatgggt 360

tgtacgcttc agcgtgggat agcaaaaagc ttaagtcagg aaaacctagt ggagttatct 420

gatgaaaatg atgatctatg tcctgtggag tgtgtcactg agttcaagac agatgatgaa 480

ttgcttagcg ttcttgaaaa gtcgaaagaa actaattctt tggttgtggt tgatttttat 540

cgcactgcat gtgggagttg taaatacata gagcagggct tctcaaaact gtgcaagcaa 600

tctggccaag aagctcctgt tatcttcctt aagcataatg tggtagatga atatgatgaa 660

caatctgaag tcgcagaaag gctccgtatc aaggcggttc ctctcttcca cttctacaaa 720

aacggagttc tcttagaatc atttgcaact agagacaagg agaggatcga cgcagctatt 780

ctcaaatata catcctcgga atcttgaaga acattcgaca aaaccgcact ctttgggtta 840

cctttcccaa atcattgaat atgatgtaac ttatattgaa caacacagca tcatccttct 900

tcttgtaaga tgtcatactg aataagacag atgcaacatt ttgatctcaa aaaaaaaaaa 960

aaaaaaaa 968

<210> 46
<211> 810
<212> DNA
<213> Arabidopsis thaliana

<400> 46
atggctgaca aagtcgatac caagttcgtc tgggttccga tttccgtccc aagtaggacg 60

gttcaagcgc atctacagaa aaacggtttg tactttcaaa agaacgcaac ttttatttcg 120

aaagttcgat cctttgggtc catagtattg acgggccact caaccaacta cgagtgctat 180

ctttcacccc tgagttcact gaaagtgttc ctgcttggtt tgctgatgag ctattgcttg 240

tggagactca ctagtagggg aggcataaac actgttgaaa accatagatg ggtctggcac 300

aacaacggag tcaggttgtc gtttccaaga gccgagtctt ctataaacat cactatgggt 360

tgtacgcttc agcgtgggat agcaaaaagc ttaagtcagg aaaacctagt ggagttatct 420

gatgaaaatg atgatctatg tcctgtggag tgtgtcactg agttcaagac agatgatgaa 480

ttgcttagcg ttcttgaaaa gtcgaaagaa actaattctt tggttgtggt tgatttttat 540

cgcactgcat gtgggagttg taaatacata gagcagggct tctcaaaact gtgcaagcaa 600

tctggtgacc aagaagctcc tgttatcttc cttaagcata atgtggtaga tgaatatgat 660

gaacaatctg aagtcgcaga aaggctccgt atcaaggcgg ttcctctctt ccacttctac 720

aaaaacggag ttctcttaga atcatttgca actagagaca aggagaggat cgacgcagct 780

attctcaaat atacatcctc ggaatcttga 810

<210> 47
<211> 1480
<212> DNA
<213> Arabidopsis thaliana

<400> 47
atggctgaca aagtcgatac caagttcgtc tgggttccga tttccgtccc aagtaggacg 60

gttcaagcgc atctacagaa aaacggtttg tactttcaaa agaacgcaac ttttatttcg 120

aaagttcgat cctttgggtc catagtattg acgggtgagt agattacaat tctgattgct 180

ttagggtttt tgccgatttg gggatggttg atctttcagt tgaatcctta atctgatttg 240

ctgacattag aaggttcatt ccatgttttg tcttaatccc catctagttg ccgaaagggt 300

ttagattttg gggtcattgc ttgaaattta tgcttcagat gcgttgatct tgtaccgttt 360

agaactctgt gttgtactcg gattcaaagt tgttcccttt acatgtcata tctctgtttt 420

ggttattttc atgcctcacc attgaattct tcgtttcagg ccactcaacc aactacgagt 480

gctatctttc acccctgagt tcactgaaag tgttcctgct tggtttgctg atgagctatt 540

gcttgtggag ttagtttaaa tatctgcaaa tgtcaagtct tcttaatgta ctgccacata 600

tcctttatat cactattacc tatgtaactt ctaggattat ttcaatgcaa agatgacccc 660

cctgctttgg ttctgcttca aaattttgat ttttgattct aatttgcaga tatctcactg 720

tagctatcat ggatactcag gactcactag taggggaggc ataaacactg ttgaaaacca 780

tagatgggtc tggcacaaca acggagtcag gttgtcgttt ccaagagccg agtcttctat 840

aaacatcact atgggttgta cgcttcagcg tgggatagca aaaagcttaa gtcaggaaaa 900

cctagtggag ttatctgatg aaaatgatga tctatgtcct gtggagtgtg tcactgagtt 960

caagacagat gatgaattgc ttagcgttct tgaaaagtcg aaagaaacta attctttggt 1020

tgtggttgat ttttatcgca ctgcatgtgg gagttgtaaa tacatagagc agggcttctc 1080

aaaactgtgc aagcaatctg gccaagaagc tcctgttatc ttccttaagc ataatgtgag 1140

tttctgatct ctcatctttg atgtataaac aataaagaag tagagaaaaa atccattaat 1200

agctctaatg ctgctatttg caggtggtag atgaatatga tgaacaatct gaagtcgcag 1260

aaaggctccg tatcaaggta atccctcttt agaattttgg tactaacaac actcaaaaat 1320

taactgaaca gcgtagtaaa gagtttgata tactacaatc ttacaggcgg ttcctctctt 1380

ccacttctac aaaaacggag ttctcttaga atcatttgca actagagaca aggagaggat 1440

cgacgcagct attctcaaat atacatcctc ggaatcttga 1480

<210> 48
<211> 1483
<212> DNA
<213> Arabidopsis thaliana

<400> 48
atggctgaca aagtcgatac caagttcgtc tgggttccga tttccgtccc aagtaggacg 60

gttcaagcgc atctacagaa aaacggtttg tactttcaaa agaacgcaac ttttatttcg 120

aaagttcgat cctttgggtc catagtattg acgggtgagt agattacaat tctgattgct 180

ttagggtttt tgccgatttg gggatggttg atctttcagt tgaatcctta atctgatttg 240

ctgacattag aaggttcatt ccatgttttg tcttaatccc catctagttg ccgaaagggt 300

ttagattttg gggtcattgc ttgaaattta tgcttcagat gcgttgatct tgtaccgttt 360

agaactctgt gttgtactcg gattcaaagt tgttcccttt acatgtcata tctctgtttt 420

ggttattttc atgcctcacc attgaattct tcgtttcagg ccactcaacc aactacgagt 480

gctatctttc acccctgagt tcactgaaag tgttcctgct tggtttgctg atgagctatt 540

gcttgtggag ttagtttaaa tatctgcaaa tgtcaagtct tcttaatgta ctgccacata 600

tcctttatat cactattacc tatgtaactt ctaggattat ttcaatgcaa agatgacccc 660

cctgctttgg ttctgcttca aaattttgat ttttgattct aatttgcaga tatctcactg 720

tagctatcat ggatactcag gactcactag taggggaggc ataaacactg ttgaaaacca 780

tagatgggtc tggcacaaca acggagtcag gttgtcgttt ccaagagccg agtcttctat 840

aaacatcact atgggttgta cgcttcagcg tgggatagca aaaagcttaa gtcaggaaaa 900

cctagtggag ttatctgatg aaaatgatga tctatgtcct gtggagtgtg tcactgagtt 960

caagacagat gatgaattgc ttagcgttct tgaaaagtcg aaagaaacta attctttggt 1020

tgtggttgat ttttatcgca ctgcatgtgg gagttgtaaa tacatagagc agggcttctc 1080

aaaactgtgc aagcaatctg gtgaccaaga agctcctgtt atcttcctta agcataatgt 1140

gagtttctga tctctcatct ttgatgtata aacaataaag aagtagagaa aaaatccatt 1200

aatagctcta atgctgctat ttgcaggtgg tagatgaata tgatgaacaa tctgaagtcg 1260

cagaaaggct ccgtatcaag gtaatccctc tttagaattt tggtactaac aacactcaaa 1320

aattaactga acagcgtagt aaagagtttg atatactaca atcttacagg cggttcctct 1380

cttccacttc tacaaaaacg gagttctctt agaatcattt gcaactagag acaaggagag 1440

gatcgacgca gctattctca aatatacatc ctcggaatct tga 1483

<210> 49
<211> 463
<212> PRT
<213> Arabidopsis thaliana

<400> 49

Met Gly Glu Glu Leu Gln Tyr Gln Gln Asn Gly Ser Ser Val Thr Ser
1 5 10 15

Asn Tyr Gly Tyr Gly Ser Ser Leu Ala Tyr Asp Gln Ser Gln Asp Ile
20 25 30

Glu Ser Ala Ala Asn Thr Ala Leu Leu Arg Glu Gln Glu Ile Glu Thr
35 40 45

Gln Lys Ile Ile Gln Gly Gln Arg Glu Ala Gly Thr Ser Val Ala Gly
50 55 60

Asp Ser Lys His Asn Thr Asp Ile Leu Arg Asp Arg Ala Asp Pro Asn
65 70 75 80

```
Ala Leu Lys Glu His Leu Leu Lys Phe Thr Ala Asn His Arg Ala Glu
                85                  90                  95

Ala Ala Ala Lys Arg Gly Gly Ser Val Ser Thr Cys Gly Glu Gly Asn
            100                 105                 110

Val Asp Val Gly Asn Gly Tyr Gly Ile Pro Gly Gly Val Ala Tyr Ala
        115                 120                 125

Gly His Ser Glu Leu Ser Gly Lys Pro Glu Pro Thr Asn Ala Ser Asn
    130                 135                 140

Asn Leu Pro Glu Tyr Leu Lys Gln Lys Leu Lys Ala Arg Gly Ile Leu
145                 150                 155                 160

Arg Asp Gly Ala Gly Ala Val Thr Ser Asn Pro Glu Asp Thr Ser Ala
                165                 170                 175

Val Ser Trp Asn Arg Gln Ala Thr Leu Pro Phe Gln Ala Asn Ala Ser
            180                 185                 190

Thr Leu Pro Leu Gly Trp Val Asp Ala Lys Asp Pro Ala Ser Gly Ala
        195                 200                 205

Thr Tyr Tyr Tyr Asn Gln His Thr Gly Thr Cys Gln Trp Glu Arg Pro
    210                 215                 220

Val Glu Leu Ser Tyr Ala Thr Ser Ser Ala Pro Pro Val Leu Ser Lys
225                 230                 235                 240

Glu Glu Trp Ile Glu Thr Phe Asp Glu Ala Ser Gly His Lys Tyr Phe
                245                 250                 255

Tyr Asn Thr Arg Thr His Val Ser Gln Trp Glu Pro Pro Ala Ser Leu
            260                 265                 270

Gln Lys Pro Ala Ala Thr Asn Ser Asn Asn Ala Val Thr Gln Ser Thr
        275                 280                 285

Ala Asn Gly Lys Gly Glu His Pro Pro Ser Gln Leu Pro Arg Cys Ser
    290                 295                 300

Gly Cys Gly Gly Trp Gly Val Gly Leu Val Gln Arg Trp Gly Tyr Cys
305                 310                 315                 320

Val His Cys Thr Arg Val Phe Asn Leu Pro Glu Lys Gln Phe Leu Pro
                325                 330                 335
```

Ala His Leu Asn His Phe Thr Asn Ala Gly Asp Ser Gly Gln Lys Asp
340 345 350

Pro Asn Gln Arg Ser Ser Ser Lys Pro Pro Met Lys Lys Val Ile Gly
355 360 365

Lys Lys Arg Ala His Ala Glu Asp Asp Glu Leu Asp Pro Met Asp Pro
370 375 380

Ser Ser Tyr Ser Asp Ala Pro Arg Gly Gly Trp Val Val Gly Leu Lys
385 390 395 400

Gly Val Gln Pro Arg Ala Ala Asp Thr Thr Ala Ser Gly Pro Leu Phe
405 410 415

Gln Gln Arg Pro Tyr Pro Ser Pro Gly Ala Val Leu Arg Arg Asn Ala
420 425 430

Glu Val Ala Ser Ser Gln Lys Lys Lys Pro Asn Ser Gln Phe Thr Glu
435 440 445

Ile Thr Lys Arg Gly Asp Gly Ser Asp Gly Leu Gly Asp Ala Asp
450 455 460

<210> 50
<211> 1648
<212> PRT
<213> Arabidopsis thaliana

<400> 50

Met Gly Ile Pro Asp Gly Ser Leu Leu Asp Leu Ile Asp Lys Val Arg
1 5 10 15

Ser Trp Ile Thr Ser Asp Ser Ser Asp Ser Leu Phe Leu Leu Ser Ser
20 25 30

Ser Lys Gln Asp Phe Gly Ile Met Pro Ile Val Ser Lys Met Cys His
35 40 45

Asp Cys Gly Thr Lys Val Glu Gln Gly Tyr Cys Cys Leu Ser Cys Gly
50 55 60

Ser Cys Trp Cys Lys Ser Cys Ser Asp Thr Glu Glu Ser Lys Met Lys
65 70 75 80

Leu Cys Arg Glu Cys Asp Ala Glu Val Arg Glu Leu Arg Val Lys Ser
85 90 95

Tyr Asp Lys Val His Pro Arg Asp Ser Pro Asp Pro Pro Ser Ser Leu
100 105 110

Ala Thr Glu Ser Glu Ser Leu Ala Ser Ser Leu Glu Ile Arg Asp Cys
115 120 125

Arg Asn Met Ala Ser Ile Arg Cys Tyr Pro Ser Arg Gly Glu Glu Glu
130 135 140

Glu Ala Arg Tyr Cys Gly Lys Gln Leu Leu Ser Pro Ser Ser Asp Asn
145 150 155 160

Tyr Gln Asp Ser Ser Asp Ile Glu Ser Gly Ser Val Ser Ala Arg His
165 170 175

Glu Leu Phe Ser Cys Lys Ser Ser Ala Gly Ser Ser Pro His Asp Ser
180 185 190

Pro Leu Arg Asn Asn Phe Ser Pro Leu Gly Arg Phe Val Gln His Ala
195 200 205

Lys Asp Leu Arg Ser Pro Thr Val Cys Ser Phe Asp Asn His Gln Glu
210 215 220

Gln Leu Leu Ala Asp Asn Leu Val Lys Pro Gly Gln Gly Val Leu Glu
225 230 235 240

Gln Glu Asp His Glu Glu Glu Glu Asp Lys Leu Gln Gln Pro Leu Asp
245 250 255

Phe Glu Asn Asn Gly Arg Ile Trp Tyr Pro Pro Pro Pro Glu Asp Glu
260 265 270

Asn Asp Asp Ala Glu Ser Asn Tyr Phe His Tyr Asp Asp Glu Asp Asp
275 280 285

Asp Ile Gly Asp Ser Ala Thr Glu Phe Ser Leu Ser Ser Ser Phe Ser
290 295 300

Ser His Ile Pro Thr Lys Glu Lys Leu Gly Glu Asn Ser Asn Glu Pro
305 310 315 320

Leu Arg Thr Val Val His Asp His Phe Arg Ala Leu Val Ala Glu Leu
325 330 335

Leu Arg Gly Glu Glu Leu Ser Pro Ser Asp Asp Gly Ser Ala Gly Glu
340 345 350

Trp Leu Asp Ile Val Thr Ala Leu Ala Trp Gln Ala Ala Asn Phe Val
355 360 365

Lys Pro Asp Thr Arg Ala Gly Gly Ser Met Asp Pro Gly Asn Tyr Val
370 375 380

Lys Ile Lys Cys Val Ala Ser Gly Asn Gln Asn Glu Ser Ile Leu Ile
385 390 395 400

Arg Gly Ile Val Cys Ser Lys Asn Ile Thr His Lys Arg Met Ile Ser
405 410 415

Gln Tyr Lys Asn Pro Arg Val Met Leu Leu Ala Gly Ser Leu Glu Tyr
420 425 430

Gln Arg Val Ala Gly Gln Leu Ala Ser Phe Asn Thr Leu Leu Gln Gln
435 440 445

Glu Asn Glu His Met Lys Ala Ile Ile Ala Lys Ile Glu Ser Leu Arg
450 455 460

Pro Asn Val Leu Leu Val Glu Lys Ser Ala Ser Ser Tyr Ala Gln Gln
465 470 475 480

Tyr Leu Leu Glu Lys Glu Ile Ser Leu Val Leu Asn Val Lys Arg Ser
485 490 495

Leu Leu Asp Arg Ile Ala Arg Cys Thr Gly Ala Val Leu Cys Pro Ser
500 505 510

Leu Asp Ser Ile Ser Thr Ala Arg Leu Gly His Cys Glu Leu Phe Arg
515 520 525

Thr Glu Arg Val Leu Glu Gln His Glu Ala Gly Asn Gln Ser Asn Arg
530 535 540

Lys Pro Ser Arg Thr Leu Met Tyr Phe Glu Gly Cys Pro Arg Arg Leu
545 550 555 560

Gly Cys Thr Val Val Leu Arg Gly Ser Cys Arg Glu Glu Leu Lys Lys
565 570 575

Val Lys His Val Ile Gln Tyr Ala Val Phe Ala Ala Tyr His Leu Ser
580 585 590

Leu Glu Thr Ser Phe Leu Ala Asp Glu Gly Ala Ser Leu Pro Lys Ile

595 600 605

Arg Leu Lys Gln Pro Gly Met Val Arg Thr Ala Ser Gln Arg Arg Ile
610 615 620

Ile Asp Glu Gly Ile Ser Leu Ile Thr Gln Ser Pro Thr Glu Thr Asp
625 630 635 640

Ser Gln Ala Leu Leu Glu Thr Ala Ala His Glu Asp Glu His Thr Ala
645 650 655

Pro Met Pro Glu His Glu Val Cys Glu Ser Leu Cys Glu Asp Phe Asp
660 665 670

Pro Thr Gln Ile Phe Pro Pro Ser Ser Glu Val Glu Thr Glu Gln Ser
675 680 685

Asp Thr Leu Asn Gly Asp Phe Ala Asn Asn Leu Val Thr Arg Ser Tyr
690 695 700

Ser Ser Asn Gln Leu Asn Asp Leu His Glu Pro Thr Leu Cys Leu Ser
705 710 715 720

Ser Glu Ile Pro Glu Thr Pro Thr Gln Gln Pro Ser Gly Glu Glu Asp
725 730 735

Asn Gly Arg Gly Glu Glu Glu Asn Gln Leu Val Asn Pro Gln Asp Leu
740 745 750

Pro Gln His Glu Ser Phe Tyr Glu Asp Asp Val Ser Ser Glu Tyr Phe
755 760 765

Ser Ala Ala Asp Ser His Gln Ser Ile Leu Val Ser Phe Ser Ser Arg
770 775 780

Cys Val Leu Lys Glu Ser Val Cys Glu Arg Ser Arg Leu Leu Arg Ile
785 790 795 800

Lys Phe Tyr Gly Ser Phe Asp Lys Pro Leu Gly Arg Tyr Leu Lys Asp
805 810 815

Asp Leu Phe Asp Lys Thr Ser Ser Cys Arg Ser Cys Lys Glu Leu Val
820 825 830

Asp Ala His Val Leu Cys Tyr Ser His Gln Asn Gly Asn Leu Thr Ile
835 840 845

```
Asn Val Arg Arg Leu Pro Ser Met Lys Leu Pro Gly Glu Gln Asp Gly
    850                 855                 860

Lys Ile Trp Met Trp His Arg Cys Leu Arg Cys Ala His Val Asp Gly
865                 870                 875                 880

Val Pro Pro Ala Thr Arg Arg Val Val Met Ser Asp Ala Ala Trp Gly
                885                 890                 895

Leu Ser Phe Gly Lys Phe Leu Glu Leu Ser Phe Ser Asn His Ala Thr
            900                 905                 910

Ala Asn Arg Val Ala Ser Cys Gly His Ser Leu Gln Arg Asp Cys Leu
        915                 920                 925

Arg Phe Tyr Gly Phe Gly Asn Met Val Ala Phe Phe Arg Tyr Ser Pro
    930                 935                 940

Ile Asn Ile Leu Thr Val Leu Leu Pro Pro Ser Met Leu Glu Phe Asn
945                 950                 955                 960

Ser His Pro Gln Gln Glu Trp Ile Arg Thr Glu Ala Ala Glu Leu Val
                965                 970                 975

Gly Lys Met Arg Thr Met Tyr Thr Glu Ile Ser Asp Met Leu Asn Arg
            980                 985                 990

Met Glu Glu Lys Ser Ser Leu Leu  Glu Pro Glu Gln Ser  Glu Ala Cys
        995                 1000                1005

Asp Leu  His Ser Arg Ile Ile  Gly Leu Ile Asp Gln  Leu Val Lys
    1010                1015                1020

Glu Lys  Asp Glu Tyr Asp Asp  Ala Leu Gln Pro Ile  Phe Glu Glu
    1025                1030                1035

Asn Leu  Gln Ile Gln Gly Ser  Leu Asp Ile Leu Glu  Leu Asn Arg
    1040                1045                1050

Leu Arg  Arg Ala Leu Met Ile  Gly Ala His Ala Trp  Asp His Gln
    1055                1060                1065

Leu Tyr  Leu Leu Asn Ser Gln  Leu Lys Lys Ala Ser  Val Phe Lys
    1070                1075                1080

Thr Gly  Asp Asp Asn Ala Pro  Arg Asn Pro Glu Met  His Asp Pro
    1085                1090                1095
```

Pro Lys Ile Asp Arg Arg Met Gln Glu Gly Ser Asp Glu Arg Asp
1100 1105 1110

Glu Gln Ser His Thr Asp Ser Glu Ala Asn Gly Asp Asn Lys Asp
1115 1120 1125

Pro Glu Asn Ile Pro Ser Pro Gly Thr Ser Leu Ser Glu Arg Ile
1130 1135 1140

Asp Ser Ala Trp Leu Gly Ser Phe Gln Asn Leu Glu Lys Ala Glu
1145 1150 1155

Thr Ile Ala Glu Thr Glu Gly Phe Ser Ala Val Asn Ser Ser Leu
1160 1165 1170

Arg Arg Leu Ala Arg Pro Ile Arg Val Gln Ser Phe Asp Ser Ala
1175 1180 1185

Ile Arg Phe Gln Glu Arg Ile Gln Lys Gly Leu Pro Pro Ser Ser
1190 1195 1200

Leu Tyr Leu Ser Thr Leu Arg Ser Phe His Ala Ser Gly Glu Tyr
1205 1210 1215

Arg Asn Met Val Arg Asp Pro Val Ser Asn Val Met Arg Thr Tyr
1220 1225 1230

Ser Gln Met Leu Pro Leu Glu Val Gln Lys Leu Asp Leu Ile Val
1235 1240 1245

Gly Ser Ala Pro Thr Tyr Ile Ser Ser Ala Ser Gln Met Ala Asp
1250 1255 1260

Gly Ala Arg Met Leu Ile Pro Gln Arg Gly Leu Asn Asp Ile Val
1265 1270 1275

Val Pro Val Tyr Asp Asp Asp Pro Ala Ser Val Val Ser Tyr Ala
1280 1285 1290

Ile Asn Ser Lys Glu Tyr Lys Glu Trp Ile Val Asn Lys Gly Leu
1295 1300 1305

Ala Ser Ser Ser Ser Ser Ser Asn Leu Asn Asn Arg Glu Ser Glu
1310 1315 1320

Pro Ser Ala Phe Ser Thr Trp Arg Ser Leu Ser Met Asp Val Asp
1325 1330 1335

Tyr Ile Gln His Ala Val Tyr Gly Ser Ser Gln Asp Asp Arg Lys
1340 1345 1350

Ser Pro His Leu Thr Ile Ser Phe Ser Asp Arg Ala Ser Ser Ser
1355 1360 1365

Ser Thr Ala Thr Glu Gly Lys Val Lys Phe Ser Val Thr Cys Tyr
1370 1375 1380

Phe Ala Thr Gln Phe Asp Thr Leu Arg Lys Thr Cys Cys Pro Ser
1385 1390 1395

Glu Val Asp Phe Val Arg Ser Leu Ser Arg Cys Gln Arg Trp Ser
1400 1405 1410

Ala Gln Gly Gly Lys Ser Asn Val Tyr Phe Ala Lys Ser Leu Asp
1415 1420 1425

Glu Arg Phe Ile Ile Lys Gln Val Val Lys Thr Glu Leu Asp Ser
1430 1435 1440

Phe Glu Asp Phe Ala Pro Glu Tyr Phe Lys Tyr Leu Lys Glu Ser
1445 1450 1455

Leu Ser Ser Gly Ser Pro Thr Cys Leu Ala Lys Ile Leu Gly Ile
1460 1465 1470

Tyr Gln Val Ser Ile Lys His Pro Lys Gly Gly Lys Glu Thr Lys
1475 1480 1485

Met Asp Leu Met Val Met Glu Asn Leu Phe Tyr Asn Arg Arg Ile
1490 1495 1500

Ser Arg Ile Tyr Asp Leu Lys Gly Ser Ala Arg Ser Arg Tyr Asn
1505 1510 1515

Pro Asn Thr Ser Gly Ala Asp Lys Val Leu Leu Asp Met Asn Leu
1520 1525 1530

Leu Glu Thr Leu Arg Thr Glu Pro Ile Phe Leu Gly Ser Lys Ala
1535 1540 1545

Lys Arg Ser Leu Glu Arg Ala Ile Trp Asn Asp Thr Asn Phe Leu
1550 1555 1560

Ala Ser Val Asp Val Met Asp Tyr Ser Leu Leu Val Gly Phe Asp

1565 1570 1575

Glu Glu Arg Lys Glu Leu Val Leu Gly Ile Ile Asp Phe Met Arg
1580 1585 1590

Gln Tyr Thr Trp Asp Lys His Leu Glu Thr Trp Val Lys Ala Ser
1595 1600 1605

Gly Ile Leu Gly Gly Pro Lys Asn Ala Ser Pro Thr Ile Val Ser
1610 1615 1620

Pro Lys Gln Tyr Lys Arg Arg Phe Arg Lys Ala Met Thr Thr Tyr
1625 1630 1635

Phe Leu Thr Val Pro Glu Pro Trp Thr Ser
1640 1645

<210> 51
<211> 574
<212> PRT
<213> Arabidopsis thaliana

<400> 51

Met Gly Thr Ser Ser Gly Ser Asn His Pro His Gln Met Leu Pro Pro
1 5 10 15

Arg Gln Gln Gln Arg Ser Gly Gly Gly Leu Glu Thr Ala Leu Ser Leu
20 25 30

Val Ser Ser Asp Gln Glu Pro Arg Arg Glu Ser Pro Ala Glu Ser Ala
35 40 45

Ser Ser Gln Glu Thr Trp Pro Leu Gly Asp Thr Val Ala Gly Lys Lys
50 55 60

Ser Met Ser Gln Lys Thr Glu Pro Asp Ser Met Glu Gln Thr Val Asn
65 70 75 80

Val Met His His Val Ser Asn Ala Asp Lys Val Ser Val Arg Asp Ile
85 90 95

Ala Arg Glu Arg Val Glu Leu Val Ala Glu Arg Met His Arg Leu Pro
100 105 110

Asp Glu Phe Leu Asp Glu Leu Lys Asn Gly Leu Lys Ser Ile Leu Glu
115 120 125

Gly Asn Val Ala Gln Ser Val Asp Glu Phe Met Phe Leu Gln Lys Val

130 135 140

Val Gln Ser Arg Thr Asp Leu Ser Ser Val Thr Leu Val Arg Ala His
145 150 155 160

Arg Val Gln Leu Glu Ile Leu Val Ala Ile Asn Thr Gly Ile Gln Ala
165 170 175

Phe Leu His Pro Asn Ile Ser Leu Ser Gln Pro Ser Leu Ile Glu Ile
180 185 190

Phe Val Tyr Lys Arg Cys Arg Asn Ile Ala Cys Gln Asn Gln Leu Pro
195 200 205

Ala Asp Asp Cys Tyr Cys Asp Ile Cys Thr Asn Arg Lys Gly Phe Cys
210 215 220

Asn Leu Cys Met Cys Thr Ile Cys Asn Lys Phe Asp Phe Ser Val Asn
225 230 235 240

Thr Cys Arg Trp Ile Gly Cys Asp Leu Cys Ser His Trp Thr His Thr
245 250 255

Asp Cys Ala Ile Arg Asp Gly Gln Ile Thr Thr Gly Ser Ser Ala Lys
260 265 270

Asn Asn Thr Ser Gly Pro Gly Glu Ile Val Phe Lys Cys Arg Ala Cys
275 280 285

Asn Arg Thr Ser Glu Leu Leu Gly Trp Val Lys Asp Val Phe Gln His
290 295 300

Cys Ala Pro Asn Trp Asp Arg Glu Ser Leu Met Lys Glu Leu Asp Phe
305 310 315 320

Val Ser Arg Ile Phe Arg Gly Ser Glu Asp Gln Arg Gly Arg Lys Leu
325 330 335

Phe Trp Lys Cys Glu Glu Leu Ile Asp Lys Ile Lys Gly Gly Leu Ala
340 345 350

Glu Ala Thr Ala Ala Lys Leu Ile Leu Met Phe Phe Gln Glu Ile Glu
355 360 365

Ser Asp Ser Ala Lys Ser Phe Glu Asn Gly Glu Gly Gly Arg Leu Met
370 375 380

Ala Pro Gln Asp Ala Cys Asn Arg Ile Ala Glu Val Val Gln Glu Thr
385 390 395 400

Leu Arg Lys Met Glu Ile Val Ala Glu Glu Lys Met Arg Met Phe Lys
405 410 415

Lys Ala Arg Met Ala Leu Glu Thr Cys Asp Arg Glu Leu Glu Asp Lys
420 425 430

Ala Lys Glu Val Ser Glu Leu Lys Ala Glu Arg Gln Lys Lys Lys Leu
435 440 445

Gln Ile Asp Glu Leu Glu Arg Ile Val Arg Leu Lys Gln Ala Glu Ala
450 455 460

Asp Met Phe Gln Leu Lys Ala Asn Glu Ala Lys Arg Glu Ala Asp Arg
465 470 475 480

Leu Gln Arg Ile Val Leu Ala Lys Met Asp Lys Ser Glu Glu Glu Tyr
485 490 495

Ala Ser Asn Tyr Leu Lys Gln Arg Leu Ser Glu Ala Glu Ala Glu Lys
500 505 510

Gln Tyr Leu Phe Glu Lys Ile Lys Leu Gln Glu Asn Ser Arg Val Ala
515 520 525

Ser Gln Ser Ser Gly Gly Gly Gly Asp Pro Ser Gln Val Met Met Tyr
530 535 540

Ser Lys Ile Arg Asp Leu Leu Gln Gly Tyr Asn Leu Ser Pro Lys Val
545 550 555 560

Asp Pro Gln Leu Asn Glu Arg Asn Pro Phe Arg Ser Asn Pro
565 570

<210> 52
<211> 268
<212> PRT
<213> Arabidopsis thaliana

<400> 52

Met Ala Asp Lys Val Asp Thr Lys Phe Val Trp Val Pro Ile Ser Val
1 5 10 15

Pro Ser Arg Thr Val Gln Ala His Leu Gln Lys Asn Gly Leu Tyr Phe
20 25 30

```
Gln Lys Asn Ala Thr Phe Ile Ser Lys Val Arg Ser Phe Gly Ser Ile
        35                  40                  45

Val Leu Thr Gly His Ser Thr Asn Tyr Glu Cys Tyr Leu Ser Pro Leu
    50                  55                  60

Ser Ser Leu Lys Val Phe Leu Leu Gly Leu Leu Met Ser Tyr Cys Leu
65                  70                  75                  80

Trp Arg Leu Thr Ser Arg Gly Gly Ile Asn Thr Val Glu Asn His Arg
                85                  90                  95

Trp Val Trp His Asn Asn Gly Val Arg Leu Ser Phe Pro Arg Ala Glu
            100                 105                 110

Ser Ser Ile Asn Ile Thr Met Gly Cys Thr Leu Gln Arg Gly Ile Ala
        115                 120                 125

Lys Ser Leu Ser Gln Glu Asn Leu Val Glu Leu Ser Asp Glu Asn Asp
    130                 135                 140

Asp Leu Cys Pro Val Glu Cys Val Thr Glu Phe Lys Thr Asp Asp Glu
145                 150                 155                 160

Leu Leu Ser Val Leu Glu Lys Ser Lys Glu Thr Asn Ser Leu Val Val
                165                 170                 175

Val Asp Phe Tyr Arg Thr Ala Cys Gly Ser Cys Lys Tyr Ile Glu Gln
            180                 185                 190

Gly Phe Ser Lys Leu Cys Lys Gln Ser Gly Gln Glu Ala Pro Val Ile
        195                 200                 205

Phe Leu Lys His Asn Val Val Asp Glu Tyr Asp Glu Gln Ser Glu Val
    210                 215                 220

Ala Glu Arg Leu Arg Ile Lys Ala Val Pro Leu Phe His Phe Tyr Lys
225                 230                 235                 240

Asn Gly Val Leu Leu Glu Ser Phe Ala Thr Arg Asp Lys Glu Arg Ile
                245                 250                 255

Asp Ala Ala Ile Leu Lys Tyr Thr Ser Ser Glu Ser
            260                 265
```

<210> 53
<211> 269
<212> PRT

<213> Arabidopsis thaliana

<400> 53

```
Met Ala Asp Lys Val Asp Thr Lys Phe Val Trp Val Pro Ile Ser Val
1               5                   10                  15

Pro Ser Arg Thr Val Gln Ala His Leu Gln Lys Asn Gly Leu Tyr Phe
            20                  25                  30

Gln Lys Asn Ala Thr Phe Ile Ser Lys Val Arg Ser Phe Gly Ser Ile
        35                  40                  45

Val Leu Thr Gly His Ser Thr Asn Tyr Glu Cys Tyr Leu Ser Pro Leu
    50                  55                  60

Ser Ser Leu Lys Val Phe Leu Leu Gly Leu Leu Met Ser Tyr Cys Leu
65                  70                  75                  80

Trp Arg Leu Thr Ser Arg Gly Gly Ile Asn Thr Val Glu Asn His Arg
                85                  90                  95

Trp Val Trp His Asn Asn Gly Val Arg Leu Ser Phe Pro Arg Ala Glu
            100                 105                 110

Ser Ser Ile Asn Ile Thr Met Gly Cys Thr Leu Gln Arg Gly Ile Ala
        115                 120                 125

Lys Ser Leu Ser Gln Glu Asn Leu Val Glu Leu Ser Asp Glu Asn Asp
    130                 135                 140

Asp Leu Cys Pro Val Glu Cys Val Thr Glu Phe Lys Thr Asp Asp Glu
145                 150                 155                 160

Leu Leu Ser Val Leu Glu Lys Ser Lys Glu Thr Asn Ser Leu Val Val
                165                 170                 175

Val Asp Phe Tyr Arg Thr Ala Cys Gly Ser Cys Lys Tyr Ile Glu Gln
            180                 185                 190

Gly Phe Ser Lys Leu Cys Lys Gln Ser Gly Asp Gln Glu Ala Pro Val
        195                 200                 205

Ile Phe Leu Lys His Asn Val Val Asp Glu Tyr Asp Glu Gln Ser Glu
    210                 215                 220

Val Ala Glu Arg Leu Arg Ile Lys Ala Val Pro Leu Phe His Phe Tyr
225                 230                 235                 240
```

Lys Asn Gly Val Leu Leu Glu Ser Phe Ala Thr Arg Asp Lys Glu Arg
245 250 255

Ile Asp Ala Ala Ile Leu Lys Tyr Thr Ser Ser Glu Ser
260 265

<210> 54
<211> 17
<212> PRT
<213> Artificial Sequence

<220>
<223> FKBP consensus motif

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> X = L or I or V or M or C

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X = Y or F

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> X = G or V or L

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> X = missing or any amino acid

<220>
<221> MISC_FEATURE
<222> (8)..(8)
<223> X = L or F or T

<220>
<221> MISC_FEATURE
<222> (9)..(10)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (12)..(14)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (15)..(15)
<223> X = D or E

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> X = S or T or A or E or Q or K

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> X = S or T or A or N

<400> 54

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa

<210> 55
<211> 27
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (2)..(23)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (25)..(26)
<223> X = any amino acid

<400> 55

Trp Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Trp Xaa Xaa Pro
20 25

<210> 56
<211> 33
<212> PRT
<213> Artificial Sequence

<220>
<223> non-classical C2-domain

<220>
<221> MISC_FEATURE
<222> (3)..(7)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> X = D or E

<220>
<221> MISC_FEATURE
<222> (26)..(26)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (28)..(29)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (32)..(32)
<223> X = missing or any amino acid

<400> 56

Lys Lys Xaa Xaa Xaa Xaa Xaa Asp Xaa Glu Leu Asp Pro Met Asp Pro
1 5 10 15

Ser Ser Tyr Ser Asp Ala Pro Arg Gly Xaa Trp Xaa Xaa Gly Leu Xaa
20 25 30

Lys

<210> 57
<211> 7
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X = any amino acid

<400> 57

Gly Gly Trp Xaa Val Gly Leu
1 5

<210> 58
<211> 19
<212> PRT
<213> Artificial Sequence

<220>
<223> Fab1 activation loop consensus

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> X = F or Y

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X = W or L

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> X = S or T or M

<220>
<221> MISC_FEATURE
<222> (14)..(15)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> X = I or L .

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> X = V or L

<400> 58

Thr Xaa Thr Xaa Asp Lys Lys Leu Glu Xaa Trp Val Lys Xaa Xaa Gly
1 5 10 15

Xaa Xaa Gly

<210> 59
<211> 6
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> X = F or W or Y

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> X = M or I or L or V

<400> 59

Pro Leu Gly Arg Xaa Xaa
1 5

<210> 60
<211> 7
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (2)..(3)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> X = R or K or H

<400> 60

Glu Xaa Xaa Gly Xaa Ile Trp
1 5

<210> 61
<211> 7
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (3)..(4)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (7)..(7)

<223> X = M or I or L or V

<400> 61

Asp Leu Xaa Xaa Pro Thr Xaa
1 5

<210> 62
<211> 8
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (3)..(4)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> X = any amino acid

<400> 62

Asp Asp Xaa Xaa Ser Xaa Tyr Phe
1 5

<210> 63
<211> 8
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> X = any amino acid

<400> 63

Thr Glu Xaa Ser Asp Xaa Leu Asn
1 5

<210> 64
<211> 6
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (2)..(2)

<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X = any amino acid

<400> 64

Cys Xaa Cys Xaa Ile Cys
1 5

<210> 65
<211> 9
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (5)..(6)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (8)..(8)
<223> X = M or I or L or V

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> X = M or I or L or V

<400> 65

Ala Cys Asn Arg Xaa Xaa Glu Xaa Xaa
1 5

<210> 66
<211> 6
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> X = any amino acid

<400> 66

Gln Xaa Ser Gly Gly Gly
1 5

<210> 67
<211> 8

<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> X = M or I or L or V

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X = any amino acid

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X = M or I or L or V

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> X = any amino acid

<400> 67

Xaa Asp Xaa Xaa Lys Xaa Gly Leu
1 5

<210> 68
<211> 8
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (3)..(4)
<223> X = any amino acid

<400> 68

Ser Glu Xaa Xaa Ala Glu Lys Gln
1 5

<210> 69
<211> 8
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (3)..(4)
<223> X = any amino acid

<220>
<221> MISC_FEATURE

<222> (8)..(8)
<223> X = D or E

<400> 69

Arg Leu Xaa Xaa Ala Glu Ala Xaa
1 5

<210> 70
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> thioredoxin-like domain

<220>
<221> MISC_FEATURE
<222> (2)..(3)
<223> X = any amino acid

<400> 70

Cys Xaa Xaa Cys
1

<210> 71
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 71
gggaattcat gtcgtatctt cttccacatc tgc 33

<210> 72
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 72
ggggatccaa atttcatttc actgaatcat gttcg 35

<210> 73
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 73

ataccactac aatggatg 18

<210> 74
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 74
agttgaagtg aacttgcggg 20

<210> 75
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Tag 100 epitope

<400> 75

Glu Glu Thr Ala Arg Phe Gln Pro Gly Tyr Arg Ser
1 5 10

<210> 76
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> c-myc epitope

<400> 76

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

<210> 77
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> FLAG epitope

<400> 77

Asp Tyr Lys Asp Asp Asp Lys
1 5

<210> 78
<211> 9
<212> PRT
<213> Artificial Sequence

<220>

<223> HA epitope

<400> 78

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

<210> 79
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> protein C epitope

<400> 79

Glu Asp Gln Val Asp Pro Arg Leu Ile Asp Gly Lys
1 5 10

<210> 80
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> VSV epitope

<400> 80

Tyr Thr Asp Ile Glu Met Asn Arg Leu Gly Lys
1 5 10

<210> 81
<211> 325
<212> PRT
<213> Arabidopsis thaliana

<400> 81

Met Gly Glu Glu Leu Gln Tyr Gln Gln Asn Gly Ser Ser Val Thr Ser
1 5 10 15

Asn Tyr Gly Tyr Gly Ser Ser Leu Ala Tyr Asp Gln Ser Gln Asp Ile
20 25 30

Glu Ser Ala Ala Asn Thr Ala Leu Leu Arg Glu Gln Glu Ile Glu Thr
35 40 45

Gln Lys Ile Ile Gln Gly Gln Arg Glu Ala Gly Thr Ser Val Ala Gly
50 55 60

Asp Ser Lys His Asn Thr Asp Ile Leu Arg Asp Arg Ala Asp Pro Asn
65 70 75 80

```
Ala Leu Lys Glu His Leu Leu Lys Phe Thr Ala Asn His Arg Ala Glu
                85                  90                  95

Ala Ala Ala Lys Arg Gly Gly Ser Val Ser Thr Cys Gly Glu Gly Asn
            100                 105                 110

Val Asp Val Gly Asn Gly Tyr Gly Ile Pro Gly Gly Val Ala Tyr Ala
        115                 120                 125

Gly His Ser Glu Leu Ser Gly Lys Pro Glu Pro Thr Asn Ala Ser Asn
    130                 135                 140

Asn Leu Pro Glu Tyr Leu Lys Gln Lys Leu Lys Ala Arg Gly Ile Leu
145                 150                 155                 160

Arg Asp Gly Ala Gly Ala Val Thr Ser Asn Pro Glu Asp Thr Ser Ala
                165                 170                 175

Val Ser Trp Asn Arg Gln Ala Thr Leu Pro Phe Gln Ala Asn Ala Ser
            180                 185                 190

Thr Leu Pro Leu Gly Trp Val Asp Ala Lys Asp Pro Ala Ser Gly Ala
        195                 200                 205

Thr Tyr Tyr Tyr Asn Gln His Thr Gly Thr Cys Gln Trp Glu Arg Pro
    210                 215                 220

Val Glu Leu Ser Tyr Ala Thr Ser Ser Ala Pro Pro Val Leu Ser Lys
225                 230                 235                 240

Glu Glu Trp Ile Glu Thr Phe Asp Glu Ala Ser Gly His Lys Tyr Phe
                245                 250                 255

Tyr Asn Thr Arg Thr His Val Ser Gln Trp Glu Pro Pro Ala Ser Leu
            260                 265                 270

Gln Lys Pro Ala Ala Thr Asn Ser Asn Asn Ala Val Thr Gln Ser Thr
        275                 280                 285

Ala Asn Gly Lys Gly Glu His Pro Pro Ser Gln Leu Pro Arg Cys Ser
    290                 295                 300

Gly Cys Gly Gly Trp Gly Val Gly Leu Val Gln Arg Trp Gly Tyr Cys
305                 310                 315                 320

Val His Cys Thr Arg
                325
```

<210> 82
<211> 49
<212> PRT
<213> Arabidopsis thaliana

<400> 82

Met Lys Lys Val Ile Gly Lys Lys Arg Ala His Ala Glu Asp Asp Glu
1 5 10 15

Leu Asp Pro Met Asp Pro Ser Ser Tyr Ser Asp Ala Pro Arg Gly Gly
20 25 30

Trp Val Val Gly Leu Lys Gly Val Gln Pro Arg Ala Ala Asp Thr Thr
35 40 45

Ala

<210> 83
<211> 31
<212> PRT
<213> Arabidopsis thaliana

<400> 83

Gly Pro Leu Phe Gln Gln Arg Pro Tyr Pro Ser Pro Gly Ala Val Leu
1 5 10 15

Arg Arg Asn Ala Glu Val Ala Ser Ser Gln Lys Lys Lys Pro Asn
20 25 30

<210> 84
<211> 19
<212> PRT
<213> Arabidopsis thaliana

<400> 84

Gln Phe Thr Glu Ile Thr Lys Arg Gly Asp Gly Ser Asp Gly Leu Gly
1 5 10 15

Asp Ala Asp

<210> 85
<211> 975
<212> DNA
<213> Arabidopsis thaliana

<400> 85
atgggagaag agctgcaata tcagcagaat ggttcatctg tgacatctaa ttatggttat 60

ggaagttcgc ttgcttacga tcaatctcaa gacattgaat ctgccgcaaa cactgctctt 120

ctccgagaac aagaaataga aacgcagaag attatacaag gtcaaagaga agctgggact 180

tcagtagccg gagactctaa gcataacact gatattcttc gagatcgtgc tgaccccaat 240

gctctaaagg aacatttgct aaagttcact gcaaatcata gagcagaagc agcggcaaag 300

cgtggtggga gtgtgagtac atgcggagaa ggtaatgtgg atgttggaaa cggttatggt 360

ataccgggtg gagttgcata cgcgggtcac tccgagctca gtgggaagcc tgaacctacc 420

aatgcttcaa acaatttgcc tgaatatctc aaacaaaaat taaaggccag gggaattctc 480

agagatggtg caggggctgt tacatccaac ccagaagata catcagcggt tagttggaat 540

agacaggcaa ctttaccttt tcaggctaat gcaagcacct taccattagg atgggtagat 600

gcaaaagatc ctgccagtgg tgctacctat tactataacc agcacactgg aacatgccag 660

tgggaaaggc ctgttgagct ttcttatgcc acttcaagtg caccacctgt actttctaaa 720

gaagagtgga ttgaaacatt tgatgaagca tctggccata agtatttcta caatacaagg 780

acacatgtgt ctcagtggga acctccagct tctttacaga agcccgctgc cacaaactct 840

aacaacgctg ttacccaaag tacggctaat gggaaggggg agcatcctcc atctcagctg 900

ccaagatgca gcggatgtgg gggctgggga gtgggccttg tccagagatg gggttattgt 960

gttcattgta ccagg 975

<210> 86
<211> 150
<212> DNA
<213> Arabidopsis thaliana

<400> 86
atgaaaaaag taataggcaa aaagcgtgct catgccgaag atgatgagtt ggatccaatg 60

gacccgagct cttactcaga tgctccacgg ggaggctggg ttgttggact gaaaggagta 120

caaccgcgag ccgctgatac aactgcttcg 150

<210> 87
<211> 90
<212> DNA
<213> Arabidopsis thaliana

<400> 87
ggtcctctgt ttcaacaacg accatatcca tcacctggag ctgttctgag gagaaacgca 60

gaggtggcat catcacagaa gaagaaacca 90

<210> 88
<211> 270
<212> DNA
<213> Arabidopsis thaliana

<400> 88
tctcagttca cagaaattac aaagagaggc gatggaagtg atggtcttgg agacgcagat 60

tgagtgagag cttccctcat cttataagtt ggttaccgtg tcttcttgtg ctttgtctgt 120

tgctttctcc atatatgaga tcaaggagag tagtttgttt acttttttagt gaatacaaac 180

aaaccatact tcttcatgtg aatttttttt aacaccagat tcaaagtatg tatacatata 240

attactcagt agcaagcaac ggtcttcaac 270

<210> 89
<211> 177
<212> PRT
<213> Arabidopsis thaliana

<400> 89

Met Ala Asp Lys Val Asp Thr Lys Phe Val Trp Val Pro Ile Ser Val
1 5 10 15

Pro Ser Arg Thr Val Gln Ala His Leu Gln Lys Asn Gly Leu Tyr Phe
20 25 30

Gln Lys Asn Ala Thr Phe Ile Ser Lys Val Arg Ser Phe Gly Ser Ile
35 40 45

Val Leu Thr Gly His Ser Thr Asn Tyr Glu Cys Tyr Leu Ser Pro Leu
50 55 60

Ser Ser Leu Lys Val Phe Leu Leu Gly Leu Leu Met Ser Tyr Cys Leu
65 70 75 80

Trp Arg Leu Thr Ser Arg Gly Gly Ile Asn Thr Val Glu Asn His Arg
85 90 95

Trp Val Trp His Asn Asn Gly Val Arg Leu Ser Phe Pro Arg Ala Glu
100 105 110

Ser Ser Ile Asn Ile Thr Met Gly Cys Thr Leu Gln Arg Gly Ile Ala
115 120 125

Lys Ser Leu Ser Gln Glu Asn Leu Val Glu Leu Ser Asp Glu Asn Asp
130 135 140

Asp Leu Cys Pro Val Glu Cys Val Thr Glu Phe Lys Thr Asp Asp Glu
145 150 155 160

Leu Leu Ser Val Leu Glu Lys Ser Lys Glu Thr Asn Ser Leu Val Val
165 170 175

Val

<210> 90
<211> 54
<212> PRT
<213> Arabidopsis thaliana

<400> 90

Val Asp Glu Tyr Asp Glu Gln Ser Glu Val Ala Glu Arg Leu Arg Ile
1 5 10 15

Lys Ala Val Pro Leu Phe His Phe Tyr Lys Asn Gly Val Leu Leu Glu
20 25 30

Ser Phe Ala Thr Arg Asp Lys Glu Arg Ile Asp Ala Ala Ile Leu Lys
35 40 45

Tyr Thr Ser Ser Glu Ser
50

<210> 91
<211> 531
<212> DNA
<213> Arabidopsis thaliana

<400> 91
atggctgaca aagtcgatac caagttcgtc tgggttccga tttccgtccc aagtaggacg 60

gttcaagcgc atctacagaa aaacggtttg tactttcaaa agaacgcaac ttttatttcg 120

aaagttcgat cctttgggtc catagtattg acgggccact caaccaacta cgagtgctat 180

ctttcacccc tgagttcact gaaagtgttc ctgcttggtt tgctgatgag ctattgcttg 240

tggagactca ctagtagggg aggcataaac actgttgaaa accatagatg ggtctggcac 300

aacaacggag tcaggttgtc gtttccaaga gccgagtctt ctataaacat cactatgggt 360

tgtacgcttc agcgtgggat agcaaaaagc ttaagtcagg aaaacctagt ggagttatct 420

gatgaaaatg atgatctatg tcctgtggag tgtgtcactg agttcaagac agatgatgaa 480

ttgcttagcg ttcttgaaaa gtcgaaagaa actaattctt tggttgtggt t 531

<210> 92
<211> 306
<212> DNA
<213> Arabidopsis thaliana

<400> 92
gtagatgaat atgatgaaca atctgaagtc gcagaaaggc tccgtatcaa ggcggttcct 60

ctcttccact tctacaaaaa cggagttctc ttagaatcat ttgcaactag agacaaggag 120

aggatcgacg cagctattct caaatataca tcctcggaat cttgaagaac attcgacaaa 180

accgcactct ttgggttacc tttcccaaat cattgaatat gatgtaactt atattgaaca 240

acacagcatc atccttcttc ttgtaagatg tcatactgaa taagacagat gcaacatttt 300

gatctc 306

<210> 93
<211> 165
<212> DNA
<213> Arabidopsis thaliana

<400> 93
gtagatgaat atgatgaaca atctgaagtc gcagaaaggc tccgtatcaa ggcggttcct 60

ctcttccact tctacaaaaa cggagttctc ttagaatcat ttgcaactag agacaaggag 120

aggatcgacg cagctattct caaatataca tcctcggaat cttga 165

<210> 94
<211> 880
<212> DNA
<213> Arabidopsis thaliana

<400> 94
atggctgaca aagtcgatac caagttcgtc tgggttccga tttccgtccc aagtaggacg 60

gttcaagcgc atctacagaa aaacggtttg tactttcaaa agaacgcaac ttttatttcg 120

aaagttcgat cctttgggtc catagtattg acgggccact caaccaacta cgagtgctat 180

ctttcacccc tgagttcact gaaagtgttc ctgcttggtt tgctgatgag ctattgcttg 240

tggagttagt ttaaatatct gcaaatgtca agtcttctta atatatctca ctgtagctat 300

catggatact caggactcac tagtagggga ggcataaaca ctgttgaaaa ccatagatgg 360

gtctggcaca acaacggagt caggttgtcg tttccaagag ccgagtcttc tataaacatc 420

actatgggtt gtacgcttca gcgtgggata gcaaaaagct taagtcagga aaacctagtg 480

gagttatctg atgaaaatga tgatctatgt cctgtggagt gtgtcactga gttcaagaca 540

gatgatgaat tgcttagcgt tcttgaaaag tcgaaagaaa ctaattcttt ggttgtggtt 600

gatttttatc gcactgcatg tgggagttgt aaatacatag agcagggctt ctcaaaactg 660

tgcaagcaat ctggtgacca agaagctcct gttatcttcc ttaagcataa tgtggtagat 720

gaatatgatg aacaatctga agtcgcagaa aggctccgta tcaaggcggt tcctctcttc 780

cacttctaca aaaacggagt tctcttagaa tcatttgcaa ctagagacaa ggagaggatc 840

gacgcagcta ttctcaaata tacatcctcg gaatcttgaa 880

<210> 95
<211> 204
<212> PRT
<213> Arabidopsis thaliana

<400> 95

Met Ser Ser Leu Leu Asn Ile Ser His Cys Ser Tyr His Gly Tyr Ser
1 5 10 15

Gly Leu Thr Ser Arg Gly Gly Ile Asn Thr Val Glu Asn His Arg Trp
20 25 30

Val Trp His Asn Asn Gly Val Arg Leu Ser Phe Pro Arg Ala Glu Ser
35 40 45

Ser Ile Asn Ile Thr Met Gly Cys Thr Leu Gln Arg Gly Ile Ala Lys
50 55 60

Ser Leu Ser Gln Glu Asn Leu Val Glu Leu Ser Asp Glu Asn Asp Asp
65 70 75 80

Leu Cys Pro Val Glu Cys Val Thr Glu Phe Lys Thr Asp Asp Glu Leu
85 90 95

Leu Ser Val Leu Glu Lys Ser Lys Glu Thr Asn Ser Leu Val Val Val
100 105 110

Asp Phe Tyr Arg Thr Ala Cys Gly Ser Cys Lys Tyr Ile Glu Gln Gly
115 120 125

Phe Ser Lys Leu Cys Lys Gln Ser Gly Asp Gln Glu Ala Pro Val Ile
130 135 140

Phe Leu Lys His Asn Val Val Asp Glu Tyr Asp Glu Gln Ser Glu Val
145 150 155 160

Ala Glu Arg Leu Arg Ile Lys Ala Val Pro Leu Phe His Phe Tyr Lys
165 170 175

Asn Gly Val Leu Leu Glu Ser Phe Ala Thr Arg Asp Lys Glu Arg Ile
180 185 190

Asp Ala Ala Ile Leu Lys Tyr Thr Ser Ser Glu Ser
195 200

<210> 96
<211> 37
<212> PRT
<213> Arabidopsis thaliana

<400> 96

Val Phe Asn Leu Pro Glu Lys Gln Phe Leu Pro Ala His Leu Asn His
1 5 10 15

Phe Thr Asn Ala Gly Asp Ser Gly Gln Lys Asp Pro Asn Gln Arg Ser
20 25 30

Ser Ser Lys Pro Pro
35

<210> 97
<211> 406
<212> PRT
<213> Arabidopsis thaliana

<400> 97

Lys Leu Asp Leu Ile Val Gly Ser Ala Pro Thr Tyr Ile Ser Ser Ala
1 5 10 15

Ser Gln Met Ala Asp Gly Ala Arg Met Leu Ile Pro Gln Arg Gly Leu
20 25 30

Asn Asp Ile Val Val Pro Val Tyr Asp Asp Asp Pro Ala Ser Val Val
35 40 45

Ser Tyr Ala Ile Asn Ser Lys Glu Tyr Lys Glu Trp Ile Val Asn Lys
50 55 60

Gly Leu Ala Ser Ser Ser Ser Ser Ser Asn Leu Asn Asn Arg Glu Ser
65 70 75 80

Glu Pro Ser Ala Phe Ser Thr Trp Arg Ser Leu Ser Met Asp Val Asp
85 90 95

Tyr Ile Gln His Ala Val Tyr Gly Ser Ser Gln Asp Asp Arg Lys Ser
100 105 110

Pro His Leu Thr Ile Ser Phe Ser Asp Arg Ala Ser Ser Ser Ser Thr
115 120 125

Ala Thr Glu Gly Lys Val Lys Phe Ser Val Thr Cys Tyr Phe Ala Thr
130 135 140

Gln Phe Asp Thr Leu Arg Lys Thr Cys Cys Pro Ser Glu Val Asp Phe
145 150 155 160

Val Arg Ser Leu Ser Arg Cys Gln Arg Trp Ser Ala Gln Gly Gly Lys
165 170 175

Ser Asn Val Tyr Phe Ala Lys Ser Leu Asp Glu Arg Phe Ile Ile Lys
180 185 190

```
Gln Val Val Lys Thr Glu Leu Asp Ser Phe Glu Asp Phe Ala Pro Glu
        195                 200                 205

Tyr Phe Lys Tyr Leu Lys Glu Ser Leu Ser Ser Gly Ser Pro Thr Cys
    210                 215                 220

Leu Ala Lys Ile Leu Gly Ile Tyr Gln Val Ser Ile Lys His Pro Lys
225                 230                 235                 240

Gly Gly Lys Glu Thr Lys Met Asp Leu Met Val Met Glu Asn Leu Phe
                245                 250                 255

Tyr Asn Arg Arg Ile Ser Arg Ile Tyr Asp Leu Lys Gly Ser Ala Arg
            260                 265                 270

Ser Arg Tyr Asn Pro Asn Thr Ser Gly Ala Asp Lys Val Leu Leu Asp
        275                 280                 285

Met Asn Leu Leu Glu Thr Leu Arg Thr Glu Pro Ile Phe Leu Gly Ser
    290                 295                 300

Lys Ala Lys Arg Ser Leu Glu Arg Ala Ile Trp Asn Asp Thr Asn Phe
305                 310                 315                 320

Leu Ala Ser Val Asp Val Met Asp Tyr Ser Leu Leu Val Gly Phe Asp
                325                 330                 335

Glu Glu Arg Lys Glu Leu Val Leu Gly Ile Ile Asp Phe Met Arg Gln
            340                 345                 350

Tyr Thr Trp Asp Lys His Leu Glu Thr Trp Val Lys Ala Ser Gly Ile
        355                 360                 365

Leu Gly Gly Pro Lys Asn Ala Ser Pro Thr Ile Val Ser Pro Lys Gln
    370                 375                 380

Tyr Lys Arg Arg Phe Arg Lys Ala Met Thr Thr Tyr Phe Leu Thr Val
385                 390                 395                 400

Pro Glu Pro Trp Thr Ser
                405
```

<210> 98
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 98

Arg Asp Ser Pro Asp Pro Pro Ser Ser Leu Ala Thr Glu Ser Glu Ser
1 5 10 15

Leu Ala Ser Ser Leu Glu Ile Arg
20

<210> 99
<211> 25
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 99

Lys Gln Leu Leu Ser Pro Ser Ser Asp Asn Tyr Gln Asp Ser Ser Asp
1 5 10 15

Ile Glu Ser Gly Ser Val Ser Ala Arg
20 25

<210> 100
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 100

Arg Ile Trp Tyr Pro Pro Pro Pro Glu Asp Glu Asn Asp Asp Ala Glu
1 5 10 15

Ser Asn Tyr Phe His
20

<210> 101
<211> 42
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 101

His Glu Val Cys Glu Ser Leu Cys Glu Asp Phe Asp Pro Thr Gln Ile
1 5 10 15

Phe Pro Pro Ser Ser Glu Val Glu Thr Glu Gln Ser Asp Thr Leu Asn
20 25 30

Gly Asp Phe Ala Asn Asn Leu Val Thr Arg
35 40

<210> 102
<211> 27
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 102

His Glu Pro Thr Leu Cys Leu Ser Ser Glu Ile Pro Glu Thr Pro Thr
1 5 10 15

Gln Gln Pro Ser Gly Glu Glu Asp Asn Gly Arg
20 25

<210> 103
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 103

Lys Ser Ser Leu Leu Glu Pro Glu Gln Ser Glu Ala Cys Asp Leu His
1 5 10 15

<210> 104
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 104

Lys Asp Pro Glu Asn Ile Pro Ser Pro Gly Thr Ser Leu Ser Glu Arg
1 5 10 15

<210> 105
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 105

Arg Glu Ser Glu Pro Ser Ala Phe Ser Thr Trp Arg

1 5 10

<210> 106
<211> 144
<212> PRT
<213> Arabidopsis thaliana

<400> 106

Asp Lys Ala Lys Glu Val Ser Glu Leu Lys Ala Glu Arg Gln Lys Lys
1 5 10 15

Lys Leu Gln Ile Asp Glu Leu Glu Arg Ile Val Arg Leu Lys Gln Ala
20 25 30

Glu Ala Asp Met Phe Gln Leu Lys Ala Asn Glu Ala Lys Arg Glu Ala
35 40 45

Asp Arg Leu Gln Arg Ile Val Leu Ala Lys Met Asp Lys Ser Glu Glu
50 55 60

Glu Tyr Ala Ser Asn Tyr Leu Lys Gln Arg Leu Ser Glu Ala Glu Ala
65 70 75 80

Glu Lys Gln Tyr Leu Phe Glu Lys Ile Lys Leu Gln Glu Asn Ser Arg
85 90 95

Val Ala Ser Gln Ser Ser Gly Gly Gly Gly Asp Pro Ser Gln Val Met
100 105 110

Met Tyr Ser Lys Ile Arg Asp Leu Leu Gln Gly Tyr Asn Leu Ser Pro
115 120 125

Lys Val Asp Pro Gln Leu Asn Glu Arg Asn Pro Phe Arg Ser Asn Pro
130 135 140

<210> 107
<211> 23
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 107

Arg Glu Ser Pro Ala Glu Ser Ala Ser Ser Gln Glu Thr Trp Pro Leu
1 5 10 15

Gly Asp Thr Val Ala Gly Lys
20

<210> 108
<211> 37
<212> PRT
<213> Arabidopsis thaliana

<400> 108

Asp Phe Tyr Arg Thr Ala Cys Gly Ser Cys Lys Tyr Ile Glu Gln Gly
1 5 10 15

Phe Ser Lys Leu Cys Lys Gln Ser Gly Gln Glu Ala Pro Val Ile Phe
20 25 30

Leu Lys His Asn Val
35

<210> 109
<211> 38
<212> PRT
<213> Arabidopsis thaliana

<400> 109

Asp Phe Tyr Arg Thr Ala Cys Gly Ser Cys Lys Tyr Ile Glu Gln Gly
1 5 10 15

Phe Ser Lys Leu Cys Lys Gln Ser Gly Asp Gln Glu Ala Pro Val Ile
20 25 30

Phe Leu Lys His Asn Val
35

<210> 110
<211> 28
<212> PRT
<213> Artificial Sequence

<220>
<223> pest sequence

<400> 110

Lys Ser Leu Ser Gln Glu Asn Leu Val Glu Leu Ser Asp Glu Asn Asp
1 5 10 15

Asp Leu Cys Pro Val Glu Cys Val Thr Glu Phe Lys
20 25

<210> 111
<211> 73
<212> PRT
<213> Arabidopsis thaliana

<400> 111

```
Asn Arg Lys Gly Phe Cys Asn Leu Cys Met Cys Thr Ile Cys Asn Lys
1               5                  10                  15

Phe Asp Phe Ser Val Asn Thr Cys Arg Trp Ile Gly Cys Asp Leu Cys
            20                  25                  30

Ser His Trp Thr His Thr Asp Cys Ala Ile Arg Asp Gly Gln Ile Thr
        35                  40                  45

Thr Gly Ser Ser Ala Lys Asn Asn Thr Ser Gly Pro Gly Glu Ile Val
    50                  55                  60

Phe Lys Cys Arg Ala Cys Asn Arg Thr
65                  70
```